28 July 2006 | $10

Science

HIV/AIDS

LATIN AMERICA & CARIBBEAN

AAAS

Anthony S. Fauci, M.D., is director of the National Institute of Allergy and Infectious Diseases at the National Institutes of Health, Bethesda, MD, USA. E-mail: Afauci@niaid.nih.gov

Twenty-Five Years of HIV/AIDS

ON 5 JUNE 1981, A REPORT IN THE *MORBIDITY AND MORTALITY WEEKLY REPORT* (*MMWR*) described five young and previously healthy gay men with *Pneumocystis carinii* pneumonia (PCP) in Los Angeles. One month later, a second report in *MMWR* described 26 men in New York and California with Kaposi's sarcoma and 10 more PCP cases in California. No one who read those reports, certainly not this author, could have imagined that this was the first glimpse of a historic era in the annals of global health.

Twenty-five years later, the human immunodeficiency virus (HIV), the cause of acquired immunodeficiency syndrome (AIDS), has reached virtually every corner of the globe, infecting more than 65 million people. Of these, 25 million have died.

The resources devoted to AIDS research over the past quarter-century have been unprecedented; $30 billion has been spent by the U.S. National Institutes of Health (NIH) alone. Investigators throughout the world rapidly discovered the etiologic agent and established the direct relationship between HIV and AIDS, developed a blood test, and delineated important aspects of HIV pathogenesis, natural history, and epidemiology. Treatment was initially confined to palliative care and management of opportunistic infections, but soon grew to include an arsenal of antiretroviral drugs (ARVs). These drugs have dramatically reduced HIV-related morbidity and mortality wherever they have been deployed. The risk factors associated with HIV transmission have been well defined. Even without a vaccine, HIV remains an entirely preventable disease in adults; and behavior modification, condom use, and other approaches have slowed HIV incidence in many rich countries and a growing number of poor ones.

With most pathogens, this narrative would sound like an unqualified success story. Yet it is very clear that scientific advances, although necessary for the ultimate control of HIV/AIDS, are not sufficient. Many important challenges remain, and in several of these the global effort is failing. New infections in 2005 still outstripped deaths by 4.1 to 2.8 million: The pandemic continues to expand. Despite substantial progress, only 20% of individuals in low- and middle-income countries who need ARVs are receiving them. Worldwide, fewer than one in five people who are at risk of becoming infected with HIV has access to basic prevention services, which even when available are confounded by complex societal and cultural issues. Stigma and discrimination associated with HIV/AIDS, and sometimes community or even governmental denial of the disease, too often dissuade individuals from getting tested or receiving medical care. Women's rights remain elusive at best in many cultures. Worldwide, thousands of women and girls are infected with HIV daily in settings where saying no to sex or insisting on condom use is not an option because of cultural factors, lack of financial independence, and even the threat of violence.

In the laboratory and the clinic, HIV continues to resist our efforts to find a cure (eradication of the virus from an infected individual) or a vaccine. In 25 years, there has not been a single well-documented report of a person whose immune system has completely cleared the virus, with or without the help of ARVs. This is a formidable obstacle to the development of an effective vaccine, for we will need to do better than nature rather than merely mimic natural infection, an approach that has worked well with many other microbes. The development of next-generation therapies and prevention tools, including topical microbicides that can empower women to directly protect themselves, will require a robust and sustained commitment to funding the best science.

Meanwhile, as we enter the second quarter-century of AIDS, we know that existing HIV treatments and prevention modalities, when appropriately applied, can be enormously effective. Programs such as President Bush's Emergency Plan for AIDS Relief; the Global Fund to Fight AIDS, Tuberculosis, and Malaria; and the efforts of philanthropies and nongovernmental organizations have clearly shown that HIV services can indeed be delivered in the poorest of settings, despite prior skepticism. We cannot lose sight of the fact that these programs must be sustained. As we commemorate the first 25 years of HIV/AIDS and celebrate our many successes, we are sobered by the enormous challenges that remain. Let us not forget that history will judge us as a global society by how well we address the next 25 years of HIV/AIDS as much as by what we have done in the first 25 years.

– Anthony S. Fauci

10.1126/science.1131993

CREDITS (TOP TO BOTTOM): NIH; MALCOLM LINTON

Published by AAAS

Re-engineering rice
423

Venomous breakdown
427

BIOMEDICAL RESEARCH

States, Foundations Lead the Way After Bush Vetoes Stem Cell Bill

Last week was a roller-coaster ride for supporters of legislation to make more human embryonic stem (ES) cell lines available to federally funded researchers. After achieving a long-sought victory in the Senate, the bill, H.R. 810, fell to a presidential veto on 19 July.

But to many, George W. Bush's action only marked another step into an era in which private entities and state governments assume greater responsibility for the funding of biomedical research. Rather than being despondent over the veto, many stem cell advocates are feeling pumped up. One is California Governor Arnold Schwarzenegger, who announced last week that the state is loaning the California Institute of Regenerative Medicine (CIRM) $150 million to get rolling. "I think with one stroke, the president energized the CIRM program," said CIRM President Zach Hall at a 20 July press conference. Sean Morrison, a stem cell researcher at the University of Michigan, Ann Arbor, agrees that the president's veto speech was "the best advertising we could have asked for." In fact, he says, a donor handed university officials a check for $50,000 right after the White House announcement.

Schwarzenegger's action, in effect, buys up most of the $200 million in "bond anticipation notes" that the state treasurer arranged for last year as a "bridge loan" while CIRM awaits the resolution of lawsuits that have obstructed the $3 billion bond issue voters passed in November 2004. CIRM board Chair Robert Klein has already gotten commitments for most of the remaining $50 million. Hall said the new money will go for research grants, with checks going out early next year.

Schwarzenegger, a Republican, was not the only governor to respond quickly to the Bush veto. Illinois Democrat Rod Blagojevich, who wants state legislators to approve $100 million for a stem cell program, announced that he is diverting $5 million from his budget for the research on top of $10 million awarded to seven Illinois institutions earlier this year. Other states, including Maryland, Massachusetts, and New Jersey, are eager to become hotbeds of stem cell research, and Missouri is poised to enter the fray should voters this fall approve an amendment to the state constitution that would legalize human ES cell research.

Not Waiting for Uncle Sam

Nonfederal funders of research on human embryonic stem cells include:

State commitments:

California:	$3 billion over 10 years
Connecticut:	$100 million over 10 years
Illinois:	$15 million via executive order
Maryland:	$15 million this year, as a start
New Jersey:	$5 million this year
Wisconsin:	$5 million to attract companies

Upcoming ballot issues:

Florida	2008
Missouri	2006
New Jersey	2006

Private donations that include support for ES cells:

Donor	Amount	Recipient
Michael Bloomberg	$100 million	Johns Hopkins U.
Starr Foundation	$50 million	Rockefeller U., Cornell U., MSKCC
Broad Foundation	$25 million	U. Southern California
Ray and Dagmar Dolby	$16 million	U. California, San Francisco
Sue and William Goss	$10 million	U. California, Irvine
Stowers Medical Institute	$10 million	Kevin Eggan and Chad Cowan, Harvard U.
Leon D. Black	$10 million	Mount Sinai School of Medicine
Private individuals	nearly $40 million	Harvard Stem Cell Institute

A yes vote in Missouri—polls show the initiative leading by 2 to 1—would unleash the Stowers Institute for Medical Research in Kansas City. The 6-year-old Stowers, with an endowment of $2.5 billion, is keen to fund human ES cell research but has been restricted by strong right-to-life forces in the state. Recently, Stowers circumvented the problem by setting up a Stowers Medical Institute in Cambridge, Massachusetts, which is supporting Harvard stem cell researcher Kevin Eggan to the tune of $6 million over 5 years. Another Harvard researcher, Chad Cowan, was recently added to the Stowers payroll. The institute is now awaiting the result of the ballot initiative. Stowers President William Neaves says the institute plans to "aggressively recruit" top stem cell researchers, as many as it can get, over the next 2 years. If the initiative passes, they will work in Missouri; if not, Stowers intends to establish new programs in stem-cell-friendly states.

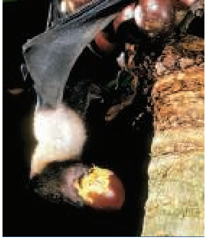
The nation's largest private medical philanthropy, the Howard Hughes Medical Institute (HHMI), is also likely to be funding more stem cell research. Although HHMI doesn't target particular research areas, its president, Thomas Cech, says that "nature abhors [the] vacuum" created by National Institutes of Health funding restrictions. He says 26 of the institute's 310 investigators "have said they plan to use human ES cells at some point"—in addition to eight who already do so.

Another private entity planning an expanded role is the Broad Foundation in Los Angeles, California, which has already donated $25 million for a center at the University of Southern California in Los Angeles. "We're looking at what else is happening at UCLA [the University of California, Los Angeles] and elsewhere," says Eli Broad. "If they can't get other funding for facilities or programs, we'll look at making grants." As for the presidential veto, he, too, says, "I think it will stimulate more private participation."

Stem cell researcher Evan Snyder of the Burnham Institute in San Diego, California, agrees. He speculates that large foundations such as the March of Dimes and the American Heart Association (AHA) may rethink their policies. AHA, for example, funds research on adult stem cells but stays away from human

CREDITS: CHARLES DHARAPAK/AP PHOTO (IMAGE); COMPILED BY NEWS STAFF FROM MEDIA SOURCES

FOCUS

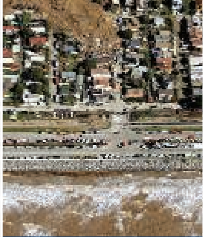


Bats, brains, and brouhaha

428

Grid computing

433



Rivers of rain

435

ES cells. Snyder also thinks venture capitalists, who have largely stayed away from human ES cells as both controversial and too far from market readiness, will be more willing to invest in the work. Currently, only two biotech companies, Geron and Advanced Cell Technology (ACT), are invested in a big way in human ES cells. "I really feel this issue has just begun in terms of public debate," says ACT CEO William Caldwell.

Indeed, a major but unquantifiable resource for stem cell research has been large gifts by private individuals. Harvard spokesperson B. D. Colen says that most of the $40 million in private funds raised by the Harvard Stem Cell Institute has come from individuals. Says Morrison: "It's not very often that an opportunity this good comes along for private philanthropy to play a leadership role in biomedical research." Access to private and state funds may also allow scientists to attempt to cultivate disease-specific cell populations through the use of somatic cell nuclear transfer. The technique, otherwise known as research cloning, would not have been permitted even under H.R. 810, and that prohibition is not expected to change in the foreseeable future.

Yet Colen and others emphasize that the federal government still plays an important role. "There's no way private philanthropy can make up for what NIH normally provides" in terms of the magnitude of funding and the chance to standardize policies and procedures, Colen says. And there's another commodity that is just as valuable as money to scientists, says Harvard stem cell researcher Len Zon: the time to pursue their research. The funding hustle "puts many researchers into a place where they're uncomfortable," says Zon. That search, he adds, "eats up time … time taken away from their research."

–CONSTANCE HOLDEN

CLIMATE CHANGE

Politicians Attack, But Evidence for Global Warming Doesn't Wilt

With hockey sticks in hand, U.S. legislators skeptical of global warming fired shots last week at what has become an iconic image in the debate. But their attack failed to change the outcome of the contest. Instead, scientists and politicians of every stripe agreed that the world is warming and that global warming is a serious issue. They also agreed to disagree about what's causing it.

On one of the hottest days of the summer in Washington, D.C., members of the investigations panel of the House Energy and Commerce Committee cast a cold eye on the so-called hockey stick curve of millennial temperature published in 1998 and 1999 papers by statistical climatologist Michael Mann of Pennsylvania State University in State College and colleagues. In a highly unusual move, the committee's chair, Representative Joe Barton (R–TX), had commissioned a statistical analysis of the contested but now-superceded curve, derived from tree rings and other proxy climate records. Statistician Edward Wegman of George Mason University in Fairfax, Virginia, Barton's choice to review Mann's work, testified that Mann's conclusion that the 1990s and 1998 were the hottest decade and year of the past millennium "cannot be supported by their analysis." An ill-advised step in Mann's statistical analysis may have created the hockey stick, Wegman said.

Because Mann wasn't there to defend himself (he was scheduled to appear at a second hearing this week), Barton bore down on the chair of a wide-ranging study of the climate of the past millennium by the U.S. National Academies' National Research Council (NRC), which also reviewed Mann's work. "No question university people like yourself believe [global warming] is caused by humans," Barton said to meteorologist Gerald North of Texas A&M University in College Station, whose 22 June NRC report concluded that the hockey stick was flawed but the sort of data on which it was based are still evidence of unprecedented warming (*Science*, 30 June, p. 1854). "My problem is that everyone seems to think we shouldn't debate the cause."

Players. Representative Joe Barton (*left*) squared off last week with Gerald North over the cause of global warming.

CREDITS (LEFT TO RIGHT): PABLO MARTINEZ MONSIVAIS/AP PHOTO; LAURA NORTH

North deflected the charge like an all-star hockey goalie. He said he doesn't disagree with Wegman's main finding that a single year or a single decade cannot be shown to be the warmest of the millennium. But that's only part of the story, he added. Finding flaws "doesn't mean Mann *et al.*'s claims are wrong," he told Barton. The recent warming may well be unprecedented, he noted, and therefore more likely to be human-induced. The claims "are just not convincing by themselves," he said. "We bring in other evidence."

The additional data include a half-dozen other reconstructions of temperatures during the past millennium. None is convincing on its own, North testified, but "our reservations should not undermine the fact that the climate is warming and will continue to warm under human influence."

North got some unexpected support from Wegman, his putative opponent on the ice. With a couple of qualifiers, Wegman agreed with North that most climate scientists have concluded that much of global warming is human-induced. And North's 12-person committee agreed with Wegman's three-person panel that the record is too fragmentary to say anything about a single year or even a single decade. The only supportable conclusion from climate proxies, the academy committee found, is that the past few decades were likely the warmest of the millennium, a conclusion of Mann's that the Wegman panel did not address. And there's a one-in-three chance that even that conclusion is wrong, North's committee found.

Consensus or not, Barton was unmoved. Scientists in the 1970s were unanimous that the next ice age was only decades away, he said. "It's the same thing" this time around, he warned.

–RICHARD A. KERR

Published by AAAS

AGRICULTURAL RESEARCH

Consortium Aims to Supercharge Rice Photosynthesis

A consortium of agricultural scientists is setting out to re-engineer photosynthesis in rice in the hope of boosting yields by 50%. It's an ambitious goal, but rice researchers say it's necessary; they seem to have hit a ceiling on rice yields, and something needs to be done to ensure a sufficient supply of the basic staple for Asia's growing population. The challenge "is very daunting, and I would say there is no certainty," says botanist Peter Mitchell of the University of Sheffield, U.K. But he adds that advances in molecular biology and genetic engineering make it a possibility.

The still-forming consortium grew out of a conference* held last week on the campus of the International Rice Research Institute (IRRI) in Los Baños, the Philippines, that drew together a small band of leading agricultural researchers from around the world. IRRI crop scientist John Sheehy says food supply and population growth in Asia are on a collision course. The Asian population is projected to increase 50% over the next 40 to 50 years, yet IRRI has not been able to increase the optimal rice yield appreciably in 30 years.

Finding a contender. An IRRI researcher measures attributes of wild rice in search of a variety suitable for supercharging.

CREDIT: IRRI

"The Green Revolution was about producing a new body for the rice plant," Sheehy says, explaining that dramatic increases in yields resulted from the introduction of semidwarf varieties that could absorb more fertilizer and take the increased weight of the grains without keeling over, a problem that plagued standard varieties. But the only answer for another dramatic increase in yields is to go under the hood of the rice plant and "supercharge" the photosynthesis engine, he says.

Evolution has provided a model of how that might be done. So-called C3 plants, such as rice, use an enzyme called RuBisCO to turn atmospheric carbon dioxide into a three-carbon compound as the first step in the carbon fixation that produces the plant's biomass. Unfortunately, RuBisCO also captures oxygen, which the plant must then shed through photorespiration, a process that causes the loss of some of the recently fixed carbon.

C4 plants, such as maize, have an additional enzyme called PEP carboxylase that initially produces a four-carbon compound that is subsequently pumped at high concentrations into cells, where it is refixed by RuBisCO. This additional step elevates the concentration of carbon dioxide around RuBisCO, crowding oxygen out and suppressing photorespiration. Consequently, C4 plants are 50% more efficient at turning solar radiation into biomass. Sheehy says theoretical predictions and some experiments at IRRI indicate that a C4 rice plant could boost potential rice yields by 50% while using less water and fertilizer.

Participants at the conference outlined a number of ways rice could be turned into a C4 plant. Evolutionary plant biologists have concluded that C4 plants evolved from C3 plants several different times. C3 plants also contain genes active in C4 plants and exhibit some aspects of the C4 cycle. Sheehy says IRRI is in the process of screening the 6000 wild rice varieties in its seed bank for wild types that may already have taken evolutionary steps toward becoming C4 plants. These might form the basis of a breeding program that could be supplemented by genes transferred from maize or other C4 plants.

Sheehy says participants at the meeting were "very optimistic" and hope that the 10 research groups in the nascent consortium will be able to demonstrate that creating C4 rice is a real possibility by 2010. If they are convinced they can make it work, they will then turn to international donors for development funding, a process that could take 12 years and cost $50 million. If C4 rice doesn't work, Asia may be heading for catastrophe. "There is no other way that has been proposed that can increase rice yields by 50%," Sheehy says.

–DENNIS NORMILE

* "Supercharging the Rice Engine," 17–21 July, IRRI, Los Baños, the Philippines.

SCIENCESCOPE

Cell Funding Stemmed

The European Union will tighten its rules over stem cell research that can be funded through its E.U.-wide research program.

In June, the E.U. Parliament voted to allow research using human embryonic stem cells in the upcoming 7-year research plan (*Science*, 23 June, p. 1732), raising hopes among stem cell scientists. But on Monday, a late-forming coalition of science ministers from countries opposed to the research threatened to block the entire program unless funding was restricted; the ministers were unwilling to fund research prohibited within their borders. After 5 hours of debate on 24 July, ministers agreed to block funding of the derivation of new stem cell lines from embryos, although there will be no restrictions on which cell lines researchers can use once they have been derived. Research Commissioner Janez Potočnik said the move preserves the status quo, because no researchers have thus far used E.U. funding to derive new cell lines.

Austin Smith of the University of Edinburgh, U.K., who heads an E.U.–funded project on stem cells, says the decision is "a compromise one can live with. The critical thing is that there is no cutoff date" for derivation of cell lines as there is for federal funding in the United States. The $63 billion Framework 7 program is to go into effect in January if the E.U. Parliament approves the change; that body next meets in the fall.

–GRETCHEN VOGEL

Bioinsecurity

Some U.S. universities handling dangerous pathogens are beefing up their security procedures in the wake of a recent federal audit. A 30 June Health and Human Services (HHS) inspector general report found that between November 2003 and November 2004, 11 of 15 universities audited lacked adequate security procedures for handling select agents. Most problems involved access control, security plans, and training. In comments on HHS's draft report, the Centers for Disease Control and Prevention stated that the findings "generally agree" with the results of its own inspections and that half of 26 identified "weaknesses" have already been addressed.

Meanwhile, Tufts University has bolstered safety steps after a test tube of botulism toxin in a centrifuge cracked at the veterinary school on 5 April. No one was hurt, but the Occupational Safety and Health Administration cited the school earlier this month for having inadequate respirators and training, fining the university $5625.

–JOCELYN KAISER

Published by AAAS

WATER PROJECTS

U.S. Senate Calls for External Reviews of Big Federal Digs

For 15 years, the U.S. Army Corps of Engineers has been locked in a battle over a $265 million project to make the Delaware River more accessible to larger ships. The corps, citing three favorable internal reviews, argues that the project is environmentally and economically sound, but opponents claim it would be bad for nearby wetlands—and would lose money. In 2002, the opponents gained some powerful ammunition from a study by the Government Accountability Office (GAO), which called the planning process for the project "fraught with errors, mistakes, and miscalculations."

Second look. Pending legislation would require the Army Corps to get outside opinions of controversial projects, such as deepening the Delaware River with dredges.

GAO's findings on the Delaware River project—currently stalled by funding disagreements among neighboring states—demonstrate the importance of regular external reviews, say the corps' many critics. And last week, they won a victory in the U.S. Senate, where legislators voted to require the use of expert panels to evaluate the engineering analyses, economic and environmental assumptions, and other aspects of projects in the corps' $2-billion-a-year construction portfolio. The corps oversees most major U.S. construction projects having to do with flood control and navigation.

A recent spate of high-profile failures and controversies, in addition to the Delaware River project, gave the measure momentum. Investigations by the University of California, Berkeley, and the American Society of Civil Engineers into last year's failure of levees in New Orleans, Louisiana, for example, found problems with design and construction that could have been avoided. Reviews of other major projects by GAO and the National Academies' National Research Council (NRC) have uncovered technical errors, inflation of benefits, and other concerns.

The additional oversight is contained in an amendment from Senators John McCain (R–AZ) and Russell Feingold (D–WI) to the Water Resources Development Act (WRDA), a bill that authorizes financing of corps projects. It would require external review of projects that cost more than $40 million or are controversial, or at the request of a federal agency or the

▶

2007 U.S. BUDGET

NIH Prepares for Lean Budget After Senate Vote

2007 is shaping up to be another year of slim pickings for the National Institutes of Health (NIH). Last week, a Senate spending panel approved a modest 0.8% increase, to $28.6 billion, for the fiscal year starting 1 October. The committee also asks the NIH director to fund a long-term, multibillion-dollar children's health study, a project NIH had said it can no longer afford.

The Senate Appropriations Committee's figure for NIH is $201 million more than President George W. Bush requested; a House spending panel last month approved roughly the amount Bush requested (minus $100 million for the Global AIDS fund). It would give most institutes a slight boost (although less than the rate of inflation) instead of the cuts proposed in the House bill. Still, the raise is far less than biomedical researchers were expecting this spring after the Senate resolved to boost spending on health and education by $7 billion.

"It's extremely concerning," says Jon Retzlaff, director of legislative relations for the Federation of American Societies for Experimental Biology (FASEB) in Bethesda, Maryland. "We are not keeping up with the advances and opportunities that are out there." Department of Labor/Health and Human Services Subcommittee Chair Arlen Specter (R–PA) noted that NIH's budget has fallen behind the rate of inflation by $3.7 billion since 2005, adding that the 2007 funding level represents a "disintegration of the appropriate federal role in health and education programs," FASEB reports.

Advocates are also worried about the committee's call for "full and timely implementation" of the projected $3.2 billion, 30-year National Children's Study (NCS). The House bill requires the National Institute of Child Health and Human Development, which oversees the study, to find $69 million within its 2007 budget. The Senate panel's report asks the NIH director's office to fund the study and added $20 million to the president's request for that office. But it doesn't specify an amount for the study itself. "We're trying to figure out" what the Senate means, says NCS Director Peter Scheidt. The report also calls for more outside scientific review of the study.

The Senate committee is silent on NIH's policy of asking grantees to submit their accepted manuscripts to NIH's free full-text papers archive. The House bill would make submission mandatory and require that NIH post the papers within 12 months.

The $141 billion spending bill, which funds NIH's parent agency and several other Cabinet-level departments, likely won't go to the Senate floor until after the November elections. The current version includes only $5 billion of the intended $7 billion increase for social programs, with NIH receiving a small slice. "All of our efforts are going … into getting the additional $2 billion," says Retzlaff, with the hope that some would flow to NIH.

The House bill has been delayed by a provision that would raise the minimum wage. After that, both chambers will meet to reconcile their two versions of the bill.

–JENNIFER COUZIN AND JOCELYN KAISER

CREDIT: CHRIS GARDNER/AP PHOTO

governor of a state affected by an upstream project. For each review, five to nine experts would be picked by someone outside the corps but within the Secretary of the Army's office.

The panel's findings and recommendations would not be binding, but the head of the corps would be required to explain why they were ignored. And in cases that go to court, judges would be required to give equal deference to the expert panel rather than simply deferring to the corps, as is customary. "It's a stick, although not a big one," says Melissa Samet of American Rivers, an advocacy group based in Washington, D.C.

In the past, the corps has heeded some outside advice, says John Boland, a water resource economist at Johns Hopkins University in Baltimore, Maryland, who has participated in many NRC reviews of corps projects. For example, the agency revamped its restoration plans related to an expansion of locks on the Upper Mississippi River after an NRC review. But the corps rejected the major criticism that its economic analysis needed fixing, and Congress authorized the $3.7 billion project as part of the new WRDA bill.

The Senate bill (S. 728) must now be melded with one passed last year by the House of Representatives (H.R. 2864) that environmentalists view as weaker. The House version allows the chief of the corps to exempt projects from external review, does not call for judicial deference, and does not require public comments to be considered. The corps declined to comment on the pending legislation, which is expected to become law by the end of the year.

–ERIK STOKSTAD

INTELLECTUAL PROPERTY

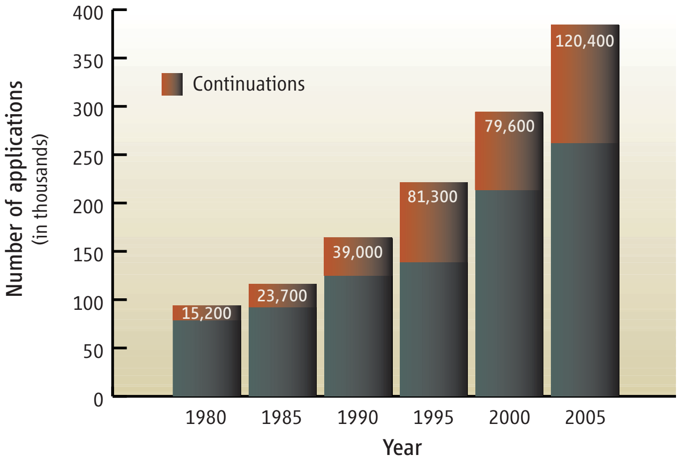
U.S. Wants to Curtail Add-On Patents to Reduce Backlog

In April 2000, Chiron Corp. received a U.S. patent for a monoclonal antibody specific to human breast cancer cells. It had actually begun the process of applying for the patent in 1984, piling on new claims even as the original application was being examined. Once the patent was awarded, Chiron sued rival California biotech Genentech, which had sold hundreds of millions of dollars of a drug, Herceptin, derived from very similar antibodies it had patented in filings made after Chiron's initial application.

Although Genentech eventually won the case, patent attorneys say that Chiron's attempt to strike back at a rival that had gotten to the market first exposes a well-used loophole in U.S. patent law: Companies can continually add detail to a pending application while benefiting from the early filing date of the initial scientific discovery. Such revised applications, known as continuations, last year made up nearly one-third of all filings with the U.S. Patent and Trademark Office (PTO).

PTO officials say the practice is drowning its workforce in paper. So in January, as part of a recent suite of reforms, the agency proposed to limit continuations to one per patent, with exceptions only on special appeals. "Examiners review the same applications over and over instead of reviewing new applications," says PTO Patent Commissioner John Doll. The new limit, he told *Science* this week, will "improve quality and move [PTO] backlog."

Although the comment period closed in May, the proposal continues to generate buzz among the intellectual-property community. Like other proposed reforms at PTO, the changes have pitted biotech companies and biomedical research institutions against the computing and software sectors. The former argue that the system works well enough now; the latter say that so-called patent trolls use continued applications to prey on true innovators. ▶



Continuous rise. A growing portion of U.S. patent applications are continuations, adding to the workload of examiners.

SOURCE: U.S. PATENT AND TRADEMARK OFFICE

SCIENCESCOPE

FDA Hunts for Conflicts ...

The U.S. Food and Drug Administration (FDA) this week announced a plan to manage conflicts of interest on its advisory committees without excluding experts with industry ties. But a key lawmaker doesn't like the idea one bit.

Under current rules, experts with industry ties can serve on FDA panels as long as they get a waiver. Legislation pending in Congress would make it tougher for FDA to appoint such experts: The House version of the law bars waivers entirely, although the Senate language is somewhat less restrictive (*Science*, 30 September 2005, p. 2145). But FDA official Scott Gottlieb, speaking at a conflict-of-interest panel this week, said that the agency "needs to preserve" the waiver system to maintain expertise. Instead, he announced that FDA will review and make more transparent its waiver-granting process.

The announcement, light on specifics, drew fire. "Saying that there are not enough potential advisory panel members available without conflicts, as the FDA argues, is an empty claim," said Representative Maurice Hinchey (D–NY) in a statement critical of FDA's plans. Hinchey is the sponsor of the House legislation. And Merrill Goozner of the Center for Science in the Public Interest, which assembled the panel, notes that some National Institutes of Health committees have instituted far stricter conflict-of-interest rules than FDA's.

–JENNIFER COUZIN

... While U.K. Slays Acronyms

The U.K. government has decided to put all of its spending on large scientific facilities in the hands of one body. The change will in effect combine the Particle Physics and Astronomy Research Council (PPARC) and the Council for the Central Laboratory of the Research Councils.

Public comments this spring ran two-to-one in favor of creating a Large Facilities Council, which would have a budget of nearly $1 billion in 2007–'08. PPARC manages the U.K. subscription to large facilities such as the CERN particle physics lab near Geneva, Switzerland, and the European Southern Observatory in Chile.

Particle physicist Brian Foster of Oxford University says he is "cautiously optimistic" about the merger but adds that PPARC had too many large commitments. So, he says, the new council's success depends on sufficient resources. Both houses of Parliament must now approve formation of the new council.

–DANIEL CLERY

Published by AAAS

A 2003 report by the Federal Trade Commission identified continuations as among the worst problems in the patent system, allowing applicants to keep patents "pending for extended periods, monitor developments in the relevant market, and then modify their claims to ensnare competitors' products." "You get to take multiple shots … and if one gets through, you're fine," says former Genentech lawyer Mark Lemley, now a law professor at Stanford University in Palo Alto, California, and an expert on continuations. The resulting uncertainty about competitors' patents, he says, "deter[s] innovation" by discouraging research investment. Semiconductor giant Micron Technology calls the reform "long overdue."

But opponents of PTO's proposed change warn that it will dampen creativity and, as California biotech Amgen noted in its public comments, "curtail the rights of true innovators to seek legitimate patent protection." Amgen officials say that biomedical research takes time and that continuations are needed to let inventors and PTO "fully understand" pending applications. Abuse is rare, they contend. The National Institutes of Health (NIH) says that continuations are needed to alert PTO to data from experiments begun before the initial application but not available for many years. (Doll says NIH could deal with such data in an appeal.)

Doll says he doesn't know when his office will issue final rules, although one of his aides told a northern Virginia audience last week that a decision is expected by January. And those rules may not be the last word. "An opportunity for a lawsuit" exists, admits Doll.

–ELI KINTISCH

IMMUNOLOGY

Mast Cells Defang Snake and Bee Venom

Venomous snakes are deadly predators; every year they kill perhaps 125,000 people, mostly in the developing world where antivenoms are less available. Researchers have long blamed immune warriors called mast cells for contributing to this toll by releasing additional toxic molecules into the victims' bodies. But a study out today puts these cells in a surprising new light.

On page 526, a team led by Stephen Galli and Martin Metz of Stanford University School of Medicine in Palo Alto, California, reports that mast cells help protect mice against snake and bee venoms, at least in part by breaking down the poisons. The "paradigm-shifting" results provide "convincing evidence for a previously unrecognized role of mast cells," says immunologist Juan Rivera of the National Institute of Arthritis and Musculoskeletal and Skin Diseases in Bethesda, Maryland.

Although mast cells help defend the body against certain parasites and bacteria, they can run amok, triggering allergic attacks including asthma and anaphylactic shock, which can be fatal. They do this by releasing molecules that induce inflammation and cause other effects that are protective in small doses but harmful if they get out of hand. These molecules include a variety of protein-splitting enzymes called proteases.

Among the proteins degraded by mast-cell proteases is endothelin-1, a potent constrictor of blood vessels that is involved in several pathological conditions including sepsis, asthma, and high blood pressure. About 2 years ago, the Galli group showed that under some circumstances this mast-cell activity protects mice against endothelin-1's toxic effects, allowing the animals to survive an infection that would otherwise throw them into septic shock.

Nearly 20 years ago, Elazar Kochva of Tel Aviv University in Israel found that the amino acid sequence of sarafotoxin, a protein in the venom of the Israeli mole viper, closely resembles that of endothelin-1. Intrigued by that similarity, Galli wondered whether mast cells protect mice against the venom. He and his colleagues tested the effects of venom provided by Kochva on normal mice and on genetically altered ones that lack mast cells. The result was clear-cut: "It takes 10 times as much venom to kill normal mice as mast cell–deficient mice," says Galli. And when mast cells derived from normal mice were engrafted into the mutant mice, the animals developed the same amount of venom resistance.

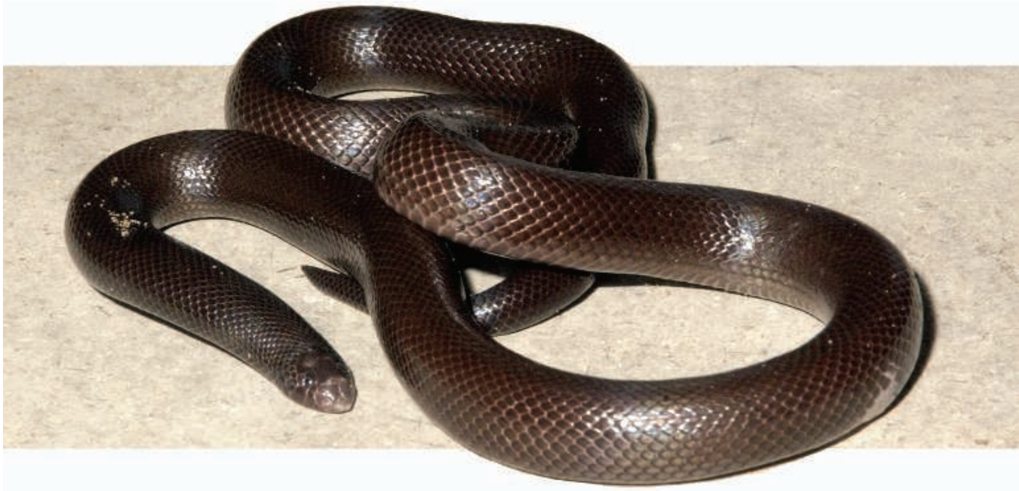
Because the Israeli mole viper lives in a limited area of the Middle East, it might be something of a biological oddity. So the Stanford team tested the venoms of the western diamondback rattlesnake and the southern copperhead, both of which are widespread in the United States. Mast cells protected mice from these venoms and also from honeybee venom. In the case of the snake venoms, Galli and his colleagues showed that a mast-cell protease called carboxypeptidase A contributes to the protection.

Hugh Miller, a mast-cell expert at the University of Edinburgh in the U.K., describes the experiments as "exceedingly elegant" demonstrations that mast cells are involved in reducing the toxic effects of venoms. Indeed, Rivera adds, "we need to rethink the role of the cells" and how they might participate in anaphylactic shock.

Both researchers caution that this mouse work doesn't prove that human mast cells also serve as an antivenom system. They point out that mouse mast cells produce more proteases than do the human versions, although both make carboxypeptidase A. Galli notes that other mast-cell products may also play a role in venom protection. One such possibility, suggested 40 years ago but not yet tested, is the anticoagulant heparin, a negatively charged molecule that might bind to, and thus inactivate, venom's positively charged components.

Given the diverse venoms that exist in nature, Galli says it's unlikely that mast cells enhance resistance to all of them. But the new work shows that the cells definitely take the bite out of some.

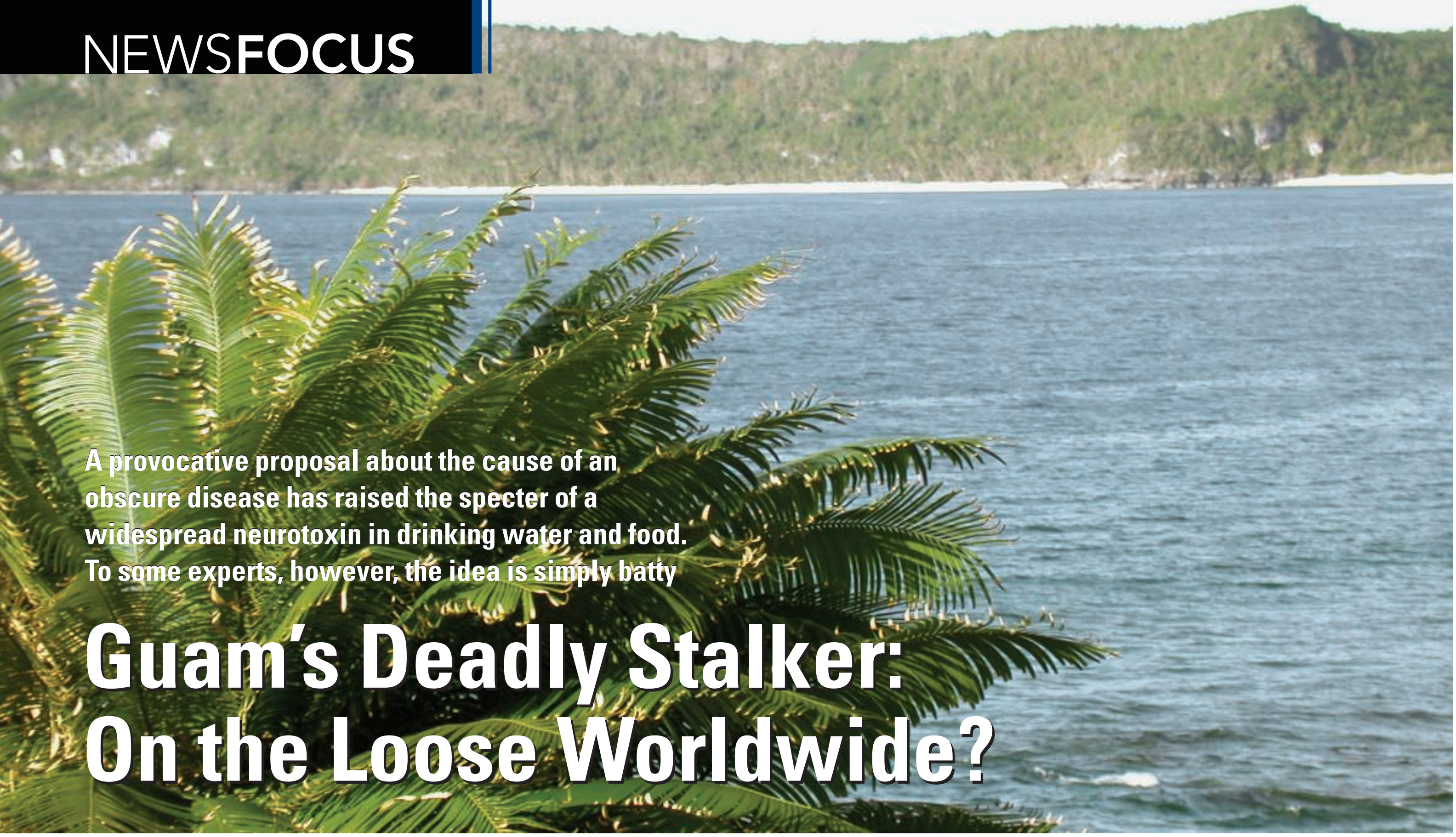
–JEAN MARX



Slithering into immunology. Venom milked from this Israeli mole viper provided the clue that led to the discovery that mast cells can protect against some snake and bee venoms.

CREDIT: M. METZ ET AL., SCIENCE

Published by AAAS



A provocative proposal about the cause of an obscure disease has raised the specter of a widespread neurotoxin in drinking water and food. To some experts, however, the idea is simply batty

Guam's Deadly Stalker: On the Loose Worldwide?

THE CASE HAS TAKEN MORE TWISTS AND turns than the most convoluted episode of the hit TV series *CSI: Crime Scene Investigation.* The killer, a fatal neurological disorder that paralyzes some victims and robs others of their minds, preyed on the Chamorro people of Guam for more than a century. Then, beginning in the 1950s, it began to retreat. Certain that something in the environment was behind the outbreak, researchers have beaten a path to the Western Pacific island in hopes that unmasking the culprit would offer clues to a mystery of profound importance: the role of environmental factors in neurodegenerative diseases around the world.

A controversial suspect emerged in 2002, when Paul Cox, an ethnobotanist then at the National Tropical Botanical Garden in Kalaheo, Hawaii, suggested that Chamorros contract the disease, which they call *lytico-bodig*, after consuming fruit bats, a traditional culinary delicacy on Guam (*Science*, 12 April 2002, p. 241). Cox and Oliver Sacks, a neurologist and popular science writer, proposed that fruit bats accumulate a toxin in their bodies from feeding on the seeds of cycads, squat, palmlike plants that thrive on Guam. Cox and colleagues have since published a string of papers supporting and extending this scenario.

The latest claim from Cox's team is even more sensational. In 2005, they reported having found the putative cycad toxin—an amino acid called β-methylamino-alanine (BMAA)—in cyanobacteria, one of the most abundant organisms on Earth. Writing in the *Proceedings of the National Academy of Sciences* (*PNAS*) last year, they proposed that BMAA could be the villain behind some of the most common neurodegenerative ailments. They argue that BMAA may find its way into drinking water and food chains and build up to neurotoxic doses in organisms at the top of the chains—such as humans.

But to many critics, cyanobacterial time bombs and fatal fruit bats smack of science fiction. "This whole thing has gotten way too far on some sloppy experimental methodology," says Daniel Perl, a neuropathologist at Mount Sinai School of Medicine in New York City who has studied *lytico-bodig* for more than 25 years. Perl and others fault Cox for making sweeping claims based on questionable samples and limited data.

Caught in the act. Fruit bats ingest a possible neurotoxin from cycad seeds.

Cox concedes that some technical concerns are valid and readily admits that his case is far from proven. "There's been some criticism, and I think that's appropriate," he says. "That's the way science works." Cox says he's determined to push forward, and some researchers argue that it's imperative his hypotheses get a fair hearing. "The implications for public health are so enormous that we have to look at this," says Deborah Mash, a neuroscientist at the University of Miami in Coral Gables, Florida, whose lab is currently probing for BMAA in the brains of North Americans who died of Alzheimer's and the muscle wasting disease amyotrophic lateral sclerosis (ALS). "If BMAA is found in ecosystems beyond Guam and we can tie it to neurodegeneration, that will be a really seminal finding," Mash says.

Links in a chain

To many scientists, *lytico-bodig* has an unquenchable allure. A solution eluded D. Carleton Gajdusek, who won half of the 1976 Nobel Prize in physiology or medicine for work on the neurodegenerative disease kuru that set the stage for the discovery of prions. Leonard Kurland, a pioneer who provided some of the first clinical descriptions of *lytico-bodig*, spent almost 50 years puzzling over the disease. Kurland "finally said to me, 'I don't care

CREDITS (TOP TO BOTTOM): THOMAS MARLER; MERLIN TUTTLE/BAT CONSERVATION INTERNATIONAL

Published by AAAS

Riddle of the tropics. Guam may hold the key to deciphering many a neurological puzzle.

who figures this out; I just want to be alive when they do,' " Perl recalls. Kurland died in December 2001.

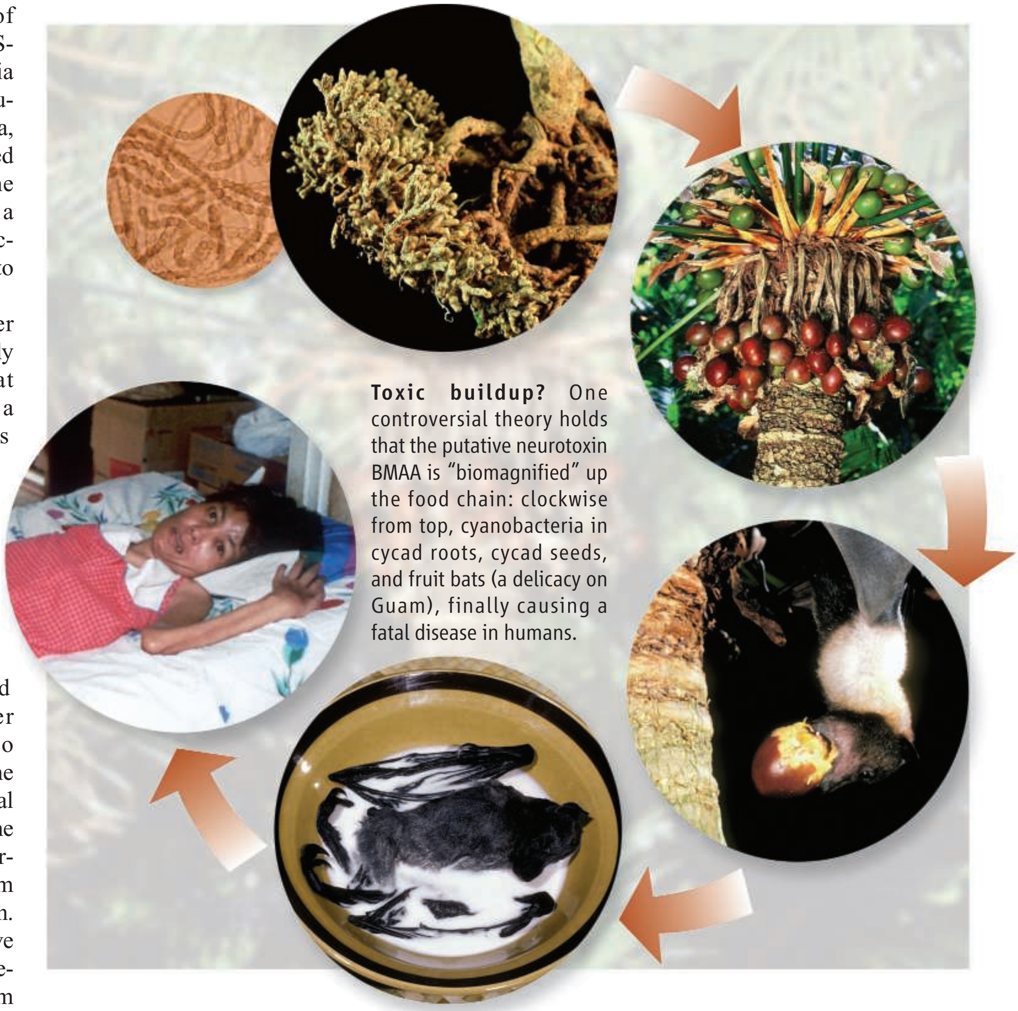
At the height of its rampage in the mid–20th century, *lytico-bodig* adopted several guises. Western experts saw a resemblance to the progressive paralysis of ALS in some cases; in others, they saw the tremors and halting movements of Parkinson's disease and the dementia of Alzheimer's. Scientists call the disorder ALS-PDC (PDC stands for Parkinsonism-dementia complex). Cases of ALS-PDC have been documented on Irian Jaya and Japan's Kii Peninsula, but most research and controversy has centered on Guam. Unmasking the cause could be the neurological equivalent of the Rosetta stone: a vital clue to deciphering the environmental factors that conspire with genetics and old age to trigger neurodegenerative illness.

Such triggers are surely out there. Fewer than 10% of Parkinson's patients have a family history of the disease, for example. What causes the remainder of Parkinson's cases is a mystery, aside from a few rare exceptions (notably, the chilling case of the "frozen addicts," a group of young drug users poisoned by a bad batch of homemade opiates in 1982). The odds of finding environmental risk factors in a large, diverse population are slim, but on Guam the small and relatively homogeneous population confines the search to a much smaller haystack.

It's hard to attribute ALS-PDC's rapid decline—from about 140 ALS cases per 100,000 people in Guam in the 1950s to fewer than 3 cases per 100,000 people in the 1990s—to anything other than an environmental cause, says Douglas Galasko, a neurologist at the University of California, San Diego, who oversees an ALS-PDC research project on Guam funded by the U.S. National Institutes of Health. "If there were a genetic cause, it wouldn't have been outbred in one generation," he says. Moreover, Chamorros who grew up outside Guam have not developed the disease, whereas some non-Chamorros who moved to the island and integrated into Chamorro society did develop it.

Suspicion fell on cycads early on. Chamorros grind the seeds to make flour for tortillas and dumplings, washing the flour several times to leach out deadly toxins. The age-old practice was observed in 1819 by the French cartographer Louis-Claude de Saulces de Freycinet. Livestock that drank from the first wash were apt to drop dead, he noted.

In the 1960s, British biochemists, trying to identify the poison, discovered BMAA; they found that it kills neurons in a petri dish. In 1987, a team led by Peter Spencer, then at Albert Einstein College of Medicine in New York City, reported in *Science* that feeding monkeys synthetic BMAA triggered neurological problems strikingly similar to ALS-PDC (*Science*, 31 July 1987, p. 517). But Gajdusek and others have argued that the findings are irrelevant to the Guam disease. They pointed out that a Chamorro would have to eat more than his own weight in cycad flour daily to get a BMAA dose equivalent to what the monkeys got. Moreover, mice given more realistic doses showed no neurodegeneration. Researchers turned to other possibilities, such as trace metals or infectious agents. But nothing definitive emerged.



Toxic buildup? One controversial theory holds that the putative neurotoxin BMAA is "biomagnified" up the food chain: clockwise from top, cyanobacteria in cycad roots, cycad seeds, and fruit bats (a delicacy on Guam), finally causing a fatal disease in humans.

Then Cox burst onto the scene. He had become interested in links between the diet and health of indigenous populations. He knew about Guam disease and that the cycad hypothesis had fallen out of favor and began to wonder whether something else in the Chamorro diet were to blame. Having previously studied the role of fruit bats as pollinators, Cox knew that hunting had helped drive one Guam species to extinction by the 1980s and another had been reduced to fewer than 100 individuals. To satisfy their taste for the furry creatures, Guamanians were importing thousands of them from Western Samoa and other islands. "I was sitting on the beach one day, and these disparate ideas came together," Cox says.

For a reality check, Cox consulted Sacks, someone he considers "sort of like Yoda," the wise Jedi Master of *Star Wars*. Sacks, who had followed the ALS-PDC saga for years, found the hypothesis intriguing, and in a 2002 paper in *Neurology*, the duo laid out the argument that a decline of native bats, known to eat cycad seeds, paralleled the disease's decline. If bats on Guam concentrate BMAA in their flesh, that could explain how humans got high enough doses to cause disease. Imported bats, on the other hand, came from islands without cycads.

To investigate the bat biomagnification hypothesis, Cox recruited one of his former graduate students, Sandra Banack, now an ecologist at California State University, Fullerton. In the August 2003 issue of *Conservation Biology*, the pair reported measurements of BMAA in cycad seeds and in the skin of three bats collected in Guam in the 1950s. These museum specimens contained hundreds of times more BMAA, gram for gram, than did the seeds. Assuming that BMAA was evenly distributed in the bats' bodies when they were alive, Cox and Banack estimated that dining on a few bats a day

CREDITS (TOP LEFT, CLOCKWISE): PAUL ALAN COX (3); MERLIN TUTTLE/BAT CONSERVATION INTERNATIONAL (2); JOHN STEELE (1)

could deliver a BMAA dose comparable to what Spencer's monkeys got.

Chamorros stew the bats with coconut milk and corn and consume them whole, says Banack, who has seen the dish prepared. These days, she says, bats are eaten at weddings and other special events. But older Chamorros have told her that when the bats were plentiful on Guam, they were more of a staple: 10 or 15 would be consumed at a single sitting. Cooking doesn't destroy BMAA.

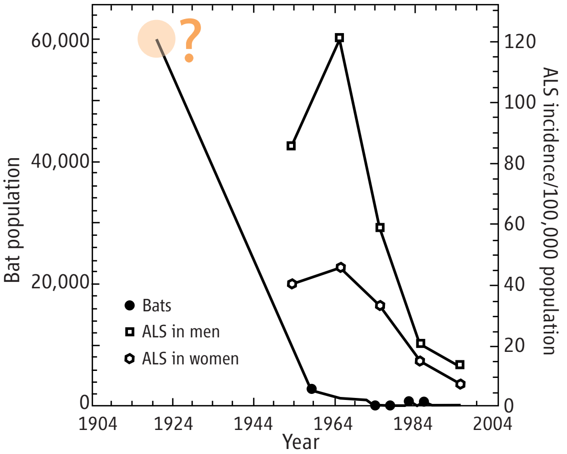
The bioaccumulation hypothesis took a twist later in 2003. Cox and Banack teamed up with Susan Murch, a plant chemist at the University of British Columbia Okanagan in Kelowna, Canada, to investigate the source of BMAA in cycad plants. Their findings pointed to nitrogen-fixing cyanobacteria. Cultured cycad roots rich with the microbes contain BMAA, whereas uninfected roots contain none, the scientists reported in *PNAS* in 2003. Free-living cyanobacteria also make BMAA, they found. Why the microbes produce the compound isn't clear, but cycads concentrate it in the outer layers of seeds, says Murch, perhaps as a defense against herbivores.

To this point, Cox's team had assembled evidence that BMAA builds up as it moves from cyanobacteria to cycads to bats. Next, the researchers looked for the compound in human brain tissue. In a 2004 paper in *Acta Neurologica Scandinavica*, they described traces of BMAA in fixed brain tissue from six Chamorros who died of ALS-PDC. The compound showed up in similar concentrations in two Canadians who died of Alzheimer's disease, but not in 13 Canadians who died of causes unrelated to neurodegenerative disease.

"We believe the people who are accumulating BMAA in North America are getting it through cyanobacteria, not cycad," Cox says. In a 2005 *PNAS* paper, he and colleagues, including cyanobacteria expert Geoffrey Codd of the University of Dundee, U.K., reported that diverse cyanobacteria—29 of 30 species tested—produce BMAA. The cyanobacteria came from soil and water samples collected in far-flung regions of the globe, which suggests that the same type of biomagnification of BMAA that Cox and his colleagues have seen on Guam may occur in other food chains. Cox says he has just begun a collaboration with Swedish scientists to investigate whether BMAA from bloom-producing cyanobacteria in the Baltic Sea accumulate in fish or other organisms.

A global danger?

At the end of 2004, Cox stepped down as director of the botanical garden to devote more time to BMAA and set up an affiliated but independently funded research facility, the Institute for Ethnomedicine in Jackson, Wyoming. "We want to test his hypothesis to see if it holds water or not," Cox says. "Quite frankly, the jury is still out."



Out of line. Critics have assailed this graph of declines in ALS-PDC rates and in Guam's fruit bats—particularly the 1920 bat population estimate.

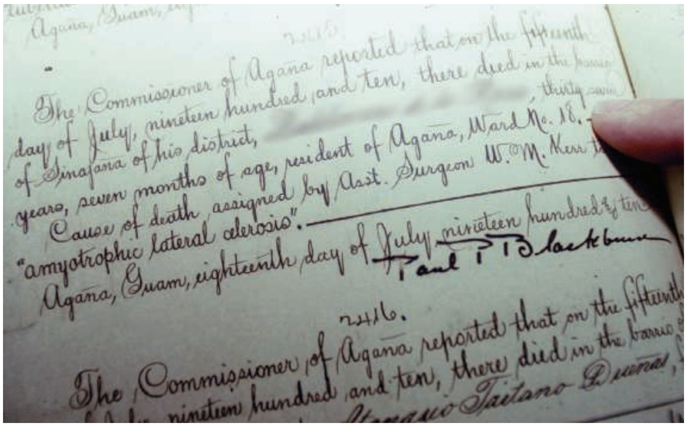
That may be an understatement. Cox's critics have assailed his hypothesis at nearly every turn, beginning with a figure in his 2002 *Neurology* paper that showed the bats on Guam and ALS-PDC incidence declining in parallel. The bat population curve is skewed by one point: a 1920s estimate of 60,000 bats on the island. In *Conservation Biology* in 2003, Cox and Banack explained that the number is derived from population estimates on nearby islands in the early 1900s combined with historical records of forest cover on Guam. Some experts say there's too much uncertainty to stake a claim on. "This is not simply sloppy science but creating data to fit the situation," asserts Anne Brooke, a wildlife biologist affiliated with U.S. Naval Base Guam and the University of Guam. Remove that point, and bat populations based on later census data taper gradually—nothing like the precipitous fall-off of ALS-PDC, she notes. "The density of bats on Guam before about 1970 is anybody's guess," Brooke says.

Because it rests on a shaky foundation, some experts insist, the bat biomagnification hypothesis is a house of cards. "They've used [the *Neurology* article] to build on all the others, referring to a correlation that in fact doesn't exist," says Christopher Shaw, a neuroscientist who studies ALS-PDC at the University of British Columbia in Vancouver, Canada. "You're allowed to speculate, but come on—don't confuse real science with imagination."

Some scientists also question the assumption that cycad seeds are a substantial part of the bats' diet. Cox and colleagues have cited a 1987 paper by wildlife biologist Gary Wiles as evidence that cycads rank among the bats' "favorite 10 food items." Wiles, now at the Washington Department of Fish and Wildlife in Olympia, had worked on Guam in the 1980s and '90s, and based on a survey of bat droppings, he compiled a list of 10 "favored" foods. Cycad seeds are on the list. However, Wiles says he never tried to quantify how much of each food the bats eat. "They've overinterpreted it," he says. "They make what I consider broad, unsubstantiated claims about the bats."

Another bone of contention is how frequently Chamorros dine on bats. "The Chamorros certainly do eat bats, but there were never enough bats for them to be a main food source," says Galasko. His team has queried islanders about their bat-eating habits. "We find no association between bat consumption and disease," he says.

Galasko and others also take issue with the Cox team's BMAA measurements. In a 2003 paper in the *Botanical Journal of the Linnean Society*, Cox and Banack reported BMAA levels based on measurements in three seeds. But Thomas Marler, a botanist at the University of Guam, has found that levels in seeds of another potential cycad toxin, sterol glucosides (see sidebar, p. 431), fluctuate according to factors such as seed age at harvest, the habitat in which seeds are collected, and how they're stored. The same would be true of BMAA or any other metabolite, Marler says. A conclusion about average BMAA concentration in cycad seeds based on just three seeds would be "more likely an artifact than reality," he contends. And that, Marler says, makes it impossible to evaluate whether BMAA levels increase from cyanobacteria to cycads to bats, as Cox and colleagues propose. In an upcoming chapter in the *Proceedings of the 2005 International Cycad Conference*, *Botanical Review*, Cox's team reports that an analysis of 52 cycad seeds of varying ages yielded an average BMAA level



Agaña, Guam,
2415
The Commissioner of Agaña reported that on the fifteenth
day of July, nineteen hundred and ten, there died in the barrio
of Sinajana of his district, [redacted], thirty-seven
years, seven months of age, resident of Agaña, Ward No. 18.
Cause of death, assigned by Asst. Surgeon W. M. Kerr
"amyotrophic lateral sclerosis".
Agaña, Guam, eighteenth day of July nineteen hundred & ten
Paul T. Blackburn

2416.
The Commissioner of Agaña reported that on the fifteenth
nineteen hundred and ten, there died in the barrio
Atanasio Taitano Dueñas,

Early evidence. Even a century ago, it was clear that Guam was struggling with an unusual plague; this 1910 death certificate notes that a 37-year-old man died of ALS.

CREDITS (TOP TO BOTTOM): ADAPTED FROM P. A. COX *ET AL.*, *NEUROLOGY* **58**, 956–959 (2002); PROVIDED BY DANIEL P. PERL

Published by AAAS

one-tenth their originally published values.

Even the evidence of BMAA in human brain tissue is under fire. Last September, neuropathologist Thomas Montine of the University of Washington, Seattle, with Galasko and Perl, failed to replicate the BMAA measurements in diseased Chamorro brains or in brains of people in the Seattle area who died of Alzheimer's disease, using high-performance liquid chromatography (HPLC). Montine suspects that the reason for the contradictory findings, reported in *Neurology* last year, may lie in differences in preservation. His group tested tissue frozen without preservatives, whereas Cox's group used tissue fixed in paraformaldehyde. Montine argues that fixed tissue should never have been used. "It does not seem to be a rigorous scientific approach to look for a methylated amino acid [BMAA] in tissue you have deliberately incubated with amino acid–modifying chemicals," he says.

Murch, the chemist who collaborated with Cox on that study, concedes that fresh brain tissue would have been better but says that the team didn't have access to such samples at the time. She counters that Montine's group used an antiquated HPLC technique that would not be sensitive enough to pick up traces of BMAA. In a letter to *Neurology* commenting on the Montine paper, Murch and others report finding BMAA in 24 frozen samples of diseased Chamorro brains—higher levels than in fixed samples from the same patients.

Even if future experiments put BMAA squarely at the crime scene—in the brains of Chamorros and others with neurodegenerative disease—the question of modus operandi remains. The evidence that BMAA is in fact a neurotoxin is mixed. Mice seem impervious. Most recently, in a paper online in *Pharmacology Biochemistry and Behavior* on 30 June, Shaw's team reports no effects in mice fed a daily BMAA dose intended to mimic levels presumably delivered by a steady diet of bats.

On the other side of the equation are Spencer's monkeys and cultured nerve cells. In a paper online in *Experimental Neurology* on 7 June, Cox, John Weiss, a neuroscientist at the University of California, Irvine, and others report that low BMAA concentrations selectively kill motor neurons in cultures of a mix of cells from mouse spinal cords. In the motor neurons, BMAA activated AMPA-kainate glutamate receptors, triggering a flood of calcium ions and boosting production of corrosive oxygen radicals.

The study hints at a possible mechanism, but researchers agree that BMAA's killer credentials will only be established with a credible animal model. "We can't claim causality until we see that lab animals fed a chronic dose develop neurological symptoms," Cox says. "That's the single biggest weakness in our idea right now."

An animal model could resolve another quandary; namely, whether BMAA kills neurons years after it's ingested. Cox and colleagues have suggested an unprecedented mechanism: BMAA, an amino acid, gets incorporated into proteins and released years later, when the proteins are broken down for recycling. In a 2004 paper in *PNAS*, Cox, Banack, and Murch describe finding protein-bound BMAA in cyanobacteria, cycad, bats, and Chamorro brain tissue. "Certainly there are people who think this is so far out," says Weiss. "My tendency is to give the exciting idea the benefit of the doubt and test it."

On Guam, meanwhile, ALS rates are now comparable to rates in the rest of the world. PDC incidence has fallen too, and it strikes people later in life. The disease seems to have transformed from one that paralyzes people in their 40s and 50s to one that causes dementia (with or without Parkinson-like rigidity) after people reach their 60s and 70s. The question, says Galasko, is "Are we simply seeing the tail end of a group of people who were exposed to something in the environment, … or are we seeing a stronger contribution from aging and genetics?" Or both?

"We haven't learned what so many of us had hoped we would learn," says John Steele, a Canadian neurologist who has worked on Guam since 1983. In his view, part of the problem is that most of the research has been done in labs far removed from Guam, the disease, and its victims. Scientists come to collect samples, he says, but rarely tarry more than a few days: "All these people who form these grand hypotheses weren't living in the midst of the disease; they were speculators at a distance." Even so, Steele says, luck has been unkind. A single clue that could break the case wide open—like the MPTP poisonings that revealed so much about Parkinson's—remains elusive. Steele once felt certain that such a break was inevitable. Now he's not sure. "I still have hope," he says. "But I no longer have confidence."

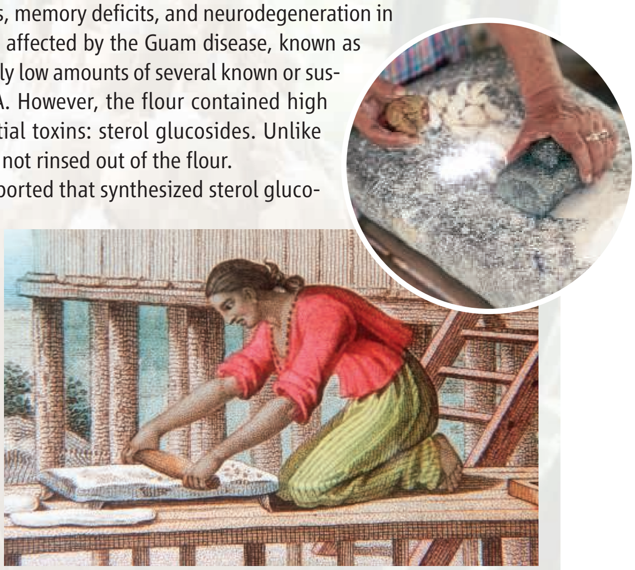
–GREG MILLER

From Cycad Flour, a New Suspect Emerges

Researchers hoping to unravel a strange neurological disorder on Guam have cast a suspicious gaze on a compound called BMAA in cycad seeds. One theory holds that fruit bats concentrate BMAA and deliver a whopping dose to anyone who eats the animals (see main text). Now, researchers led by Christopher Shaw of the University of British Columbia in Vancouver, Canada, have fingered a different suspect in cycad seeds, one that the native Chamorros of Guam ingest directly.

In 2002, Shaw, graduate student Jason Wilson, and others reported in *NeuroMolecular Medicine* that mice fed pellets of cycad flour prepared by Chamorros for their own consumption develop movement and coordination problems, memory deficits, and neurodegeneration in the spinal cord and parts of the brain affected by the Guam disease, known as ALS-PDC. Analyses revealed vanishingly low amounts of several known or suspected cycad toxins, including BMAA. However, the flour contained high amounts of another family of potential toxins: sterol glucosides. Unlike BMAA, insoluble sterol glucosides are not rinsed out of the flour.

Shaw's team has subsequently reported that synthesized sterol glucosides are lethal to cultured neurons, and at last year's meeting of the Society for Neuroscience, they described neurodegeneration in the spinal cords of mice fed sterol glucosides for up to 10 weeks. Figuring out how sterol glucosides kill neurons will be a crucial next step, Shaw says, as will looking for the compounds in ALS-PDC victims.



The old-fashioned way. Preparation of cycad flour on Guam today (*inset*) has changed little since the 19th century.

The role of sterol glucosides in neurodegenerative disease could extend far beyond Guam. "Every plant makes them," Shaw says. In a paper in press at *Medical Hypotheses*, Shaw and colleagues note that the bacterium *Helicobacter pylori* also makes compounds similar in structure to the cycad glucosides—and they point out that some studies have suggested that Parkinsonism is more common in people who have suffered gastric ulcers caused by *H. pylori*. And at the Society for Neuroscience meeting last year, Shaw's team reported having found elevated sterol glucoside levels in blood samples from 40 North American ALS patients.

Some experts are skeptical, however. Peter Spencer, a neuroscientist at Oregon Health & Science University in Portland, notes that sterol glucosides have been used in Europe to treat men with enlarged prostates—with no reported ill effects.

–G.M.

CREDITS: PROVIDED BY JOHN STEELE, FROM J. ARAGO NARRATIVE, PUBLISHED BY TREUTTEL AND WURTZ, LONDON 1823: 239–240 (AVAILABLE AT THE MARC, UNIVERSITY OF GUAM; (INSET) JOHN STEELE; (BACKGROUND IMAGE) THOMAS MARLER

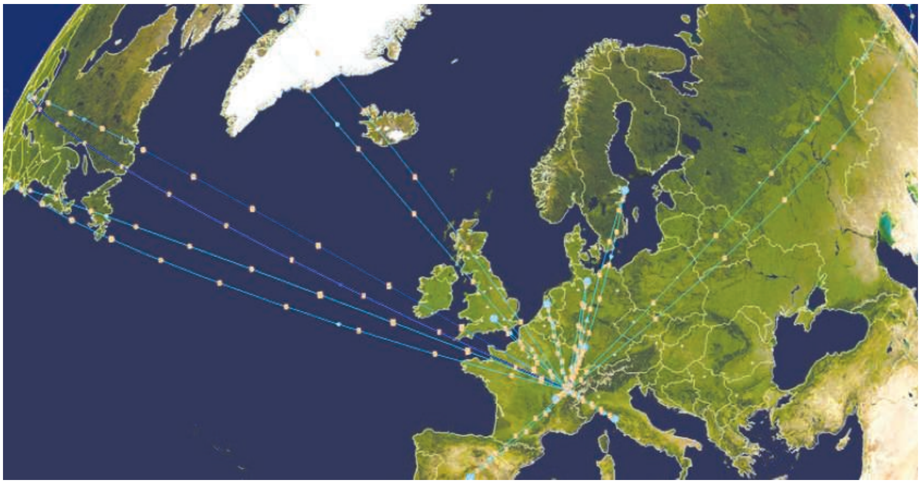
INFRASTRUCTURE

Can Grid Computing Help Us Work Together?

A different way to use the Internet aims to transform the way researchers collaborate, once the wrinkles are ironed out

Modern science is a game for collaborators. Hundreds of researchers took part in sequencing the human genome, and each of the giant detectors now being built for the Large Hadron Collider (LHC) at the CERN particle physics lab near Geneva, Switzerland, is designed and operated by teams of more than 1000 physicists and engineers. The need to work collectively and the arrival of the Internet have spawned a new style of research organization: "centers without walls," also known as virtual organizations or collaboratories.

Now, some researchers think collaboration is going to get a lot easier. For more than 10 years, groups of researchers—often allied with computer engineers and behavioral scientists—have been experimenting with new ways for widely separated teams to work together using networked computers. This process, known as cyberinfrastructure in the United States and e-science in Europe, has spawned more than just useful tools such as chatrooms and electronic blackboards; it has given birth to a whole new way of using the Internet, known as grid computing.



Practice run. CERN researchers test the speed of their grid by streaming simulated LHC data from Geneva to centers around the globe.

CREDIT: CERN

The essence of grid computing is sharing resources. A group of researchers could set up a virtual organization that shares the computer processing power in each of their institutions, as well as databases, memory storage facilities, and scientific instruments such as telescopes or particle accelerators. By pooling computer resources, anyone in the virtual organization could potentially tap into power equivalent to that of a supercomputer. "People will have to think differently about the value of collaboration," says Malcolm Atkinson, director of the e-Science Institute at the University of Edinburgh, U.K.. "Policy, culture, and behavior will all have to adapt. That's why it's not going to happen in 5 years."

As in the early days of the World Wide Web, particle physicists are leading the way. For the past 3 years, physicists have been working on an ambitious test-bed grid designed to distribute the torrents of data that will flow from LHC and allow large communities of researchers to archive, process, and study it at numerous centers around the globe. In October, the grid will be declared operational, ready for when the accelerator is completed next year. "Unless it is working, [LHC] cannot do its job. It's mission critical," says Wolfgang von Rüden, CERN's head of information technology.

Although grid computing was invented about a dozen years ago, computer experts are still struggling to make it reliable and easy to use. The difficulty lies in persuading numerous institutions—each with its own individual network architecture, firewall, and security system—to open their computing resources to outsiders. As a result, researchers still need quite a lot of computing expertise, and so uptake has been slow. But enthusiasts believe grid computing will soon reach a tipping point—as did the Internet and the World Wide Web before it—when the benefits outweigh the difficulties and no researcher can be seen without it. And if the technical hurdles can be cleared, everyone gains: Resources spend less time sitting idle and are used more efficiently. "It's not something that's going to happen overnight, but it will have a big impact," says von Rüden.

It's good to chat

An influential early attempt at computer-assisted collaboration was the Upper Atmosphere Research Collaboratory (UARC). Begun in 1992, UARC aimed to give researchers remote access to a suite of instruments operated by the U.S. National Science Foundation (NSF) at an observatory above the Arctic Circle. The instruments, including an incoherent scatter radar, observe the interaction of Earth's magnetosphere with particles streaming in from the sun. Instead of having to travel to Greenland, UARC users could gather data while sitting at their desks, annotating their observations in real time and interacting with distant colleagues using a chatroom-style interface. "It was a complex sociotechnical challenge, not just a technical one," says computer scientist Daniel Atkins, who was project director of UARC while a professor at the University of Michigan, Ann Arbor.

Later, UARC expanded to incorporate other radars around the world as well as data from research satellites. Atkins says some researchers were possessive about data at first. "But after about 5 or 6 years, they flipped around and were welcoming to others," he says. "UARC helped coalesce ideas about cyberinfrastructure."

Other collaboratories soon sprang up in disciplines as wide-ranging as earthquake engineering, nuclear fusion, biomedical informatics, and anatomy. Some computing experts began to think about using networked computers in a new way to make collaboration even easier. In 1994, Ian Foster and Steven Tuecke of Argonne National Laboratory in Illinois teamed up with Carl Kesselman of the California Institute of Technology in Pasadena to found the Globus Project, an effort to develop a software system to enable worldwide scientific cooperation. In 1997, the team released the first version of their Globus Toolkit, a set of software tools for creating grids.

Globus, and similar systems such as Condor and Moab, all work in roughly the same way. Ideally, a researcher sits down at her computer and logs into the virtual organization to which she belongs. Immediately, she can see which of her regular collaborators are online and can chat with them. She can also access the numerous archives, databases, and instruments that they share around the globe. Making use of the large combined computing power of the collaboration, she requests a computing job using an onscreen form, and then wanders off and makes coffee. A software system called middleware takes over the job and consults a catalog to see where on the grid to find the data necessary for the job and where there is available processing capacity, memory facilities for short-term storage during the job, and perhaps visualization

Published by AAAS

capacity to present the results in a way the researcher can use. Software "brokers" then manage those resources, transfer data from place to place, and monitor the progress of the job. Long before our researcher finishes her coffee, the results should be waiting for her perusal.

Particular success

In 1999, Foster and Kesselman edited a book called *The Grid: Blueprint for a New Computing Infrastructure*, which did much to popularize the idea of grid computing. CERN jumped on the bandwagon. In the 1990s, when CERN physicists were designing LHC, they soon realized that CERN's computing facilities would be swamped by the data coming from the cathedral-sized detectors they were planning to build. Les Robertson, head of the LHC Computing Grid project, says they had planned to set up a spoke-like network to channel data from CERN to a handful of large computing centers elsewhere in the world for archiving. "It was a simple model, but restrictive," Robertson says.

When CERN researchers learned about grid computing, they decided it was a better way to go. In 2003, CERN launched a test-bed grid with connections to 20 other centers. Today, it links 100 institutes worldwide and handles 25,000 jobs every day. Once LHC is operational next year, the aim is to carry out initial processing at CERN and then stream the data out to 11 "tier-1" centers where the data will be processed more intensively and archived. Particle physicists around the globe will then be able to tap into the data through the 90 or so other tier-2 centers. Much research has been done on pushing up the world speed record for distributing data over a network. "I won't claim it all works yet, but it is a useful system," Robertson says.

Although grid computing has been largely a grassroots movement, funding agencies and governments got involved once they realized it could lead to a more efficient use of computing resources and more productive collaborations. The European Union has been an enthusiastic supporter of grids, running prototypes called DataGrid and DataTag before launching the Enabling Grids for e-Science (EGEE) in April 2004. The grid now links 200 centers in 40 countries worldwide. EGEE director Robert Jones, who is based at CERN, reckons that as many as 25,000 individual computers may be connected to it. Jones says EGEE has deliberately worked to expand grid computing beyond physics. EGEE can now run applications in nine discipline areas, and there are 60 different virtual organizations using the grid.

PC farm. Quantities of off-the-shelf PCs provide cheap computer power at CERN.

In the United States, a number of discipline-specific grids supported by NSF and the Department of Energy (DOE) gradually coalesced and, in 2004, formed the Open Science Grid. "OSG came from the grassroots. It grew out of projects which decided 'Let's work together,' " says OSG Director Ruth Pordes. Some universities in the United States are also planning campuswide grids, and OSG hopes that it can eventually link up with them to expand from the 50 NSF, DOE, and university sites currently connected.

NSF also supports a number of specialized supercomputer centers, and these have clubbed together into TeraGrid. Dane Skow, TeraGrid's deputy director, explains that it is different from other grids in that the nine connected supercomputers are optimized for different jobs, such as raw number-crunching, visualization, or simulation. He sees most researchers accessing TeraGrid through discipline-specific "gateways," where they can submit a job, and then a few computer experts will work out how best to apply the job to the grid.

Perhaps the biggest impetus in the United States came from a panel chaired by Atkins that was tasked by NSF with looking at its past programs in advanced computing and seeing whether there were some new wave it should be riding. The panel consulted widely and was surprised to find scientists getting involved in the quite advanced information technology (IT) of grid computing. "We became quite excited by this science-driven, bottom-up phenomenon," says Atkins. His report, published in December 2004, advocated a new NSF program in support of cyberinfrastructure. In February, Atkins became director of NSF's new Office of Cyberinfrastructure. "There is a lot going on in [disciplinary] silos, but we need common solutions to ensure we aren't reinventing the wheel," Atkins says. "I think we will see a kind of accelerating effect over the next 5 years."

Meanwhile, developers are wrestling with the practical problems of harmonizing a tangle of incompatible networks. A body called the Global Grid Forum has been leading the effort to draw up common standards for grid computing. In June, it merged with a parallel body called the Enterprise Grid Alliance to form the Open Grid Forum. Enterprise grids work within a single company, which is easier to achieve because commercial organizations usually have a uniform network architecture and security system. The merger is "a huge step forward," says the University of Edinburgh's Atkinson.

Researchers are keen for industry to become more involved in grid computing so that, eventually, the communications industry can take it off their hands. "We're not here to do grids for the rest of our lives," says Jones. "Grid computing will only be sustainable if industry picks it up."

But some grid promoters complain that grids are taking too long to become user-friendly. "You can't give it to your mother yet. You still need to be an IT enthusiast," Jones says. "The interface needs to be improved to make it easier," says biologist Ying-Ta Wu of Academia Sinica in Taipei, who took part in an EGEE project to find possible drug components against the avian influenza virus H5N1. "We needed a lot of experts to work with." And the grids themselves still need too much hands-on maintenance to make them economical. "You still need heroes in some places," says Atkinson. "EGEE relies on many skilled and dedicated people—more than we can afford." Says Pordes: "Grids have not delivered on the original hype or promise. … [People] tried to do too much too soon."

Despite the teething troubles, many grid enthusiasts think that it is on the cusp of widespread adoption. "It has much the same feel as the early Internet," says Skow. "But there are enough usability issues to sort out that a single trigger won't push us over the top." But for Atkinson, that push is inevitable: "If this is an infection, soon it's going to turn into a pandemic."

–DANIEL CLERY

CREDIT: CERN

METEOROLOGY

Rivers in the Sky Are Flooding The World With Tropical Waters

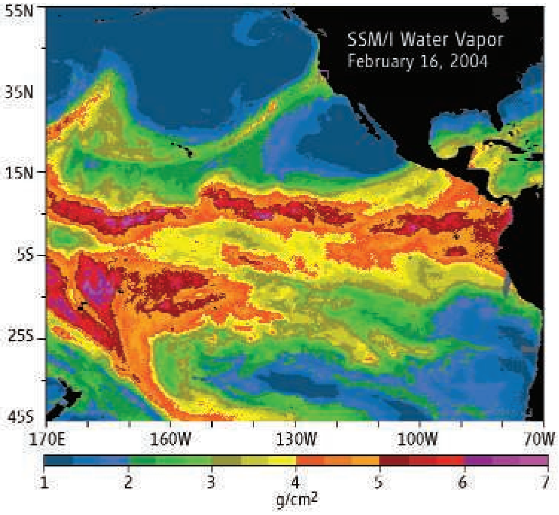
When mid-latitude storms tap into the great stores of moisture in the tropical atmosphere, the rain pours and pours, rivers rise, the land slides, and locusts can swarm

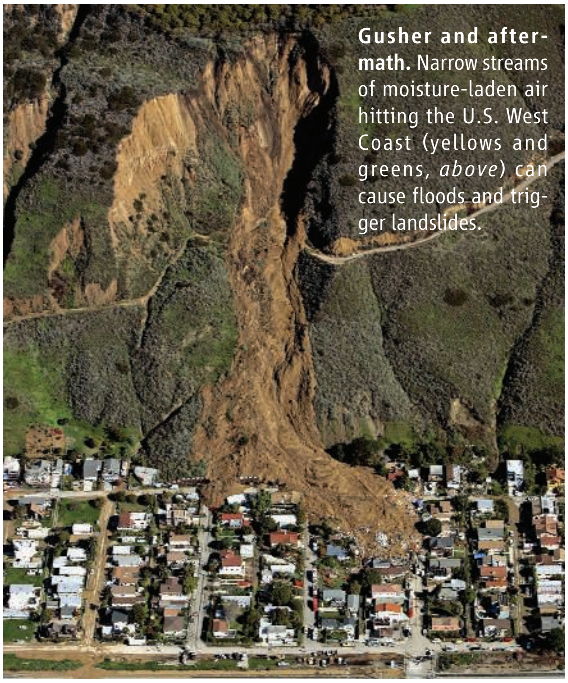
Call them tropical plumes, atmospheric rivers, Hawaiian fire hoses, or Pineapple Expresses. Whatever the label, meteorologists are now recognizing the extent to which these streams of steamy tropical air transport vast amounts of moisture across the globe, often leaving natural disasters in their wake. When a classic atmospheric river tapped tropical moisture to dump a meter of rain onto southern California in January 2005, it triggered the massive La Conchita mudslide that killed 10 people. Torrential rains fed by an atmospheric river inundated the U.S. East Coast last month, meteorologists say, and researchers recently showed that atmospheric rivers can flood places such as northwest Africa as well, with equally dramatic effects.

Researchers are now probing the workings of these rivers in the sky in hopes of forecasting them better, not only day to day but also decade to decade as the greenhouse builds. When atmospheric rivers make the connection to the moisture-laden tropics, "all hell can break loose," says meteorologist Jonathan Martin of the University of Wisconsin, Madison.

Weather forecasters have long recognized the importance of narrow streams of poleward-bound air. A glance at satellite images of the wintertime North Pacific Ocean shows great, comma-shaped storms marching eastward, their tails arcing back southwestward toward Hawaii and beyond. These storms are redressing the imbalance between the warm tropics and cold poles by creating an atmospheric conveyor belt. Cold air sweeps broadly southward behind the cold front that runs along the tail, and warm air is driven poleward along and just ahead of the front. It is this warm and inevitably moist stream paralleling the front that has come to be known as an atmospheric river.

Those storms sweeping across the mid-latitudes are obviously major conduits in the atmosphere's circulation system, but few appreciated quite how major until 1998, when meteorologists Yong Zhu and the late Reginald Newell of the Massachusetts Institute of Technology in Cambridge analyzed globe-circling weather data on winds and their water content. Although the three to five atmospheric rivers in each hemisphere at any one time occupied just 10% of the mid-latitudes, they found, the rivers were carrying fully 90% of the moisture moving poleward.





Gusher and aftermath. Narrow streams of moisture-laden air hitting the U.S. West Coast (yellows and greens, *above*) can cause floods and trigger landslides.

CREDITS (TOP TO BOTTOM): PAUL NEIMAN; KEVORK DJANSEZIAN/AP PHOTO

In 2004, meteorologist Martin Ralph of the National Oceanic and Atmospheric Administration's (NOAA's) Environmental Technology Laboratory in Boulder, Colorado, and his colleagues showed just how narrow atmospheric rivers really are. By parachuting instrument packages along a line across the cold fronts of 17 storms, they found that the core of a river—a jet of 85-kilometer-per-hour wind centered a kilometer above the surface—is something like 100 kilometers across. But the river is so moist that it moves about 50 million liters of water per second, equivalent to a 100-meter-wide pipe gushing water at 50 kilometers per hour.

Such a "fire hose of water aimed at the West Coast," as Ralph describes it, can do serious damage. Ralph and colleagues combined NOAA field studies near the coast of northern California with satellite observations in a detailed study of the February 2004 flooding of the Russian River, they reported in the 1 July *Geophysical Research Letters*. In that case, an atmospheric river extended 7000 kilometers through Hawaii, linking up with moisture-laden air from the tropics.

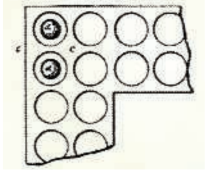
At the California coast, the mountains directed the oncoming atmospheric river upward, wringing out enough rain to create record flows on the Russian River. Near-record flows hit rivers and streams along 500 kilometers of the coast and across the breadth of California. Ralph and his colleagues also found that similar atmospheric rivers caused all seven floods on the Russian River since October 1997.

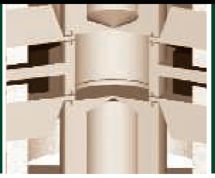
Other researchers are looking at atmospheric rivers around the world. In an upcoming paper in *Weather and Forecasting*, meteorologists Peter Knippertz of the University of Mainz, Germany, and Jonathan Martin of the University of Wisconsin, Madison, will report on an atmospheric river that dumped 8 centimeters of hail on central Los Angeles in November 2003 and went on to deliver heavy precipitation to Arizona. Last year, they described three cases on the west coast of North Africa of extremely heavy rains in 2002 and 2003 fed by atmospheric rivers. Some areas received up to a year's worth of precipitation in one storm. An autumn 2003 drenching helped create favorable breeding conditions for desert locusts, leading to devastating outbreaks in large parts of northern West Africa.

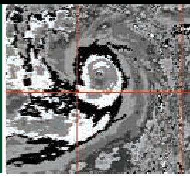
The latest studies remind meteorologists that atmospheric rivers and their flooding are commonplace. By studying them, meteorologists are hoping to improve forecasts of heavy rains and flooding; in the case of the Russian River, they expected 13 centimeters of rain, but 25 centimeters fell, setting off the record flood. Advances will come from improving the observations of atmospheric rivers offshore and correcting errors in forecast models, particularly as they simulate the encounter between atmospheric rivers and mountains. Even climate modelers hoping to predict precipitation in a greenhouse world will have to get a better handle on the rivers in the sky.

–RICHARD A. KERR

Published by AAAS







LETTERS

edited by Etta Kavanagh

Adult Stem Cell Treatments for Diseases?

OPPONENTS OF RESEARCH WITH EMBRYONIC STEM (ES) CELLS OFTEN CLAIM THAT ADULT STEM cells provide treatments for 65 human illnesses. The apparent origin of those claims is a list created by David A. Prentice, an employee of the Family Research Council who advises U.S. Senator Sam Brownback (R–KS) and other opponents of ES cell research (*1*).

Prentice has said, "Adult stem cells have now helped patients with at least 65 different human diseases. It's real help for real patients" (*2*). On 4 May, Senator Brownback stated, "I ask unanimous consent to have printed in the Record the listing of 69 different human illnesses being treated by adult and cord blood stem cells" (*3*).

In fact, adult stem cell treatments fully tested in all required phases of clinical trials and approved by the U.S. Food and Drug Administration are available to treat only nine of the conditions on the Prentice list, not 65 [or 72 (*4*)]. In particular, allogeneic stem cell therapy has proven useful in treating hematological malignancies and in ameliorating the side effects of chemotherapy and radiation. Contrary to what Prentice implies, however, most of his cited treatments remain unproven and await clinical validation. Other claims, such as those for Parkinson's or spinal cord injury, are simply untenable.

The references Prentice cites as the basis for his list include various case reports, a meeting abstract, a newspaper article, and anecdotal testimony before a Congressional committee. A review of those references reveals that Prentice not only misrepresents existing adult stem cell treatments, but also frequently distorts the nature and content of the references he cites (*5*).

> "By promoting the falsehood that adult stem cell treatments are already in general use for 65 diseases and injuries, Prentice and those who repeat **his claims mislead laypeople and cruelly deceive patients"**
>
> —Smith *et al.*

For example, to support the inclusion of Parkinson's disease on his list, Prentice cites congressional testimony by a patient (*6*) and a physician (*7*), a meeting abstract by the same physician (*8*), and two publications that have nothing to do with stem cell therapy for Parkinson's (*9*, *10*). In fact, there is currently no FDA-approved adult stem cell treatment—and no cure of any kind—for Parkinson's disease.

For spinal cord injury, Prentice cites personal opinions expressed in Congressional testimony by one physician and two patients (*11*). There is currently no FDA-approved adult stem cell treatment or cure for spinal cord injury.

The reference Prentice cites for testicular cancer on his list does not report patient response to adult stem cell therapy (*12*); it simply evaluates different methods of adult stem cell isolation.

The reference Prentice cites on non-Hodgkin's lymphoma does not assess the treatment value of adult stem cell transplantation (*13*); rather, it describes culture conditions for the laboratory growth of stem cells from lymphoma patients.

Prentice's listing of Sandhoff disease, a rare disease that affects the central nervous system, is based on a layperson's statement in a newspaper article (*14*). There is currently no cure of any kind for Sandhoff disease.

By promoting the falsehood that adult stem cell treatments are already in general use for 65 diseases and injuries, Prentice and those who repeat his claims mislead laypeople and cruelly deceive patients (*15*).

SHANE SMITH,[1] WILLIAM NEAVES,[2]* STEVEN TEITELBAUM[3]

[1]Children's Neurobiological Solutions Foundation, 1726 Franceschi Road, Santa Barbara, CA 93103, USA. [2]Stowers Institute for Medical Research, 1000 East 50th Street, Kansas City, MO 64110, USA. [3]Department of Pathology and Immunology, Washington University, 660 South Euclid Avenue, St. Louis, MO 63110, USA.

*To whom correspondence should be addressed. E-mail: William_Neaves@stowers-institute.org

References

1. Posted at the Web site of DoNoHarm, The Coalition of Americans for Research Ethics (accessed 8 May 2006 at www.stemcellresearch.org/facts/treatments.htm).
2. D. Prentice, *Christianity Today* **49** (no. 10), 71 (17 Oct. 2005) (accessed 8 May 2006 at www.christianitytoday.com/ct/2005/010/24.71.html).
3. S. Brownback, "Stem cells," *Congressional Record*, 4 May 2006 (Senate) (page S4005–S4006) (accessed 8 May 2006 at http://frwebgate6.access.gpo.gov/cgi-bin/waisgate.cgi?WAISdocID=122359256098+2+2+0&WAISaction=retrieve).
4. According the latest version of the list, accessed 12 July 2006.
5. See chart compiling and analyzing Prentice's list of 65 diseases allegedly treated by adult stem cells at the supplemental data repository available as Supporting Online Material on *Science* Online at www.sciencemag.org/cgi/content/full/1129987/DC1.
6. D. Turner, Testimony before Senator Sam Brownback's Science, Technology and Space Subcommittee on 14 July 2004 (accessed 8 May 2006 at http://commerce.senate.gov/hearings/testimony.cfm?id=1268&wit_id=3676).
7. M. Lévesque, Testimony before Senator Sam Brownback's Science, Technology and Space Subcommittee on 14 July 2004 (accessed 8 May 2006 at http://commerce.senate.gov/hearings/testimony.cfm?id=1268&wit_id=3670).
8. M. Lévesque, T. Neuman, Abstract No. 702, Annual Meeting of the American Association of Neurological Surgeons, 8 April 2002.
9. S. Gill *et al.*, *Nat. Med.* **9**, 589 (2003).
10. S. Love *et al.*, *Nat. Med.* **11**, 703 (2005).
11. M. Lévesque, Testimony before Senator Sam Brownback's Science, Technology and Space Subcommittee on 14 July 2004 (accessed 8 May 2006 at http://commerce.senate.gov/hearings/testimony.cfm?id=1268&wit_id=3670); L. Dominguez, Testimony before Senator Sam Brownback's Science, Technology and Space Subcommittee on 14 July 2004 (accessed 8 May 2006 at http://commerce.senate.gov/hearings/testimony.cfm?id=1268&wit_id=3673); S. Fajt, Testimony before Senator Sam Brownback's Science, Technology and Space Subcommittee on 14 July 2004 (accessed 8 May 2006 at http://commerce.senate.gov/hearings/testimony.cfm?id=1268&wit_id=3674).
12. K. Hanazawa *et al.*, *Int. J. Urol.* **7**, 77 (2000).
13. M. Yao *et al.*, *Bone Marrow Transpl.* **26**, 497 (2000).
14. K. Augé, "Stem cells infuse kin with hope," *Denver Post*, 24 Aug. 2004.
15. M. Enserink, *Science* **313**, 160 (2006).

Published online 13 July 2006

Published by AAAS

Name Dropping on Decapods

THE EXCITEMENT AND PUBLICITY SURROUNDING the discovery of a new and unusual decapod crustacean from Pacific hydrothermal vents ("A crustacean Yeti," Random Samples, 17 Mar., p. 1531) is well deserved. However, the new family proposed to accommodate the species is hardly "the first new family of decapods… in a century."

The most recent compilation of all currently recognized extant decapod families (*1*) lists 36 families of decapods—nearly a quarter of all recognized decapod families—that have been erected or newly recognized since 1906. Although some of the family names recognize assemblages that were previously known but only recently treated as families, many are based on novel finds. Included among these are at least two families based on species that are, like the new "Yeti crab," endemic to or restricted to hydrothermal vents and cold hydrocarbon seeps: the brachyuran crab family Bythograeidae (*2*) and the caridean shrimp family Alvinocarididae (*3*), based on the genus *Alvinocaris*, a name that honors the DSV Alvin, a submarine that was first launched in 1964.

JOEL W. MARTIN

Invertebrate Studies/Crustacea, Natural History Museum of Los Angeles County, 900 Exposition Boulevard, Los Angeles, CA 90007, USA.

References

1. J. W. Martin, G. E. Davis, An Updated Classification of the Recent Crustacea, *Nat. Hist. Mus. Los Angeles County Sci. Ser.* **39**, 1 (2001).
2. A. B. Williams, *Proc. Biol. Soc. Wash.* **93**, 443 (1980).
3. M. L. Christoffersen, *Boll. Zool. (Univ. Sao Paulo Brazil)* **10**, 273 (1986).

Questions About Mass Spectrometry Data

I AM WRITING TO EXPRESS MY PERSONAL CONcerns about Hao Xin's article "University clears Chinese biophysicist of misconduct" (News of the Week, 28 Apr., p. 511).

On 19 April, Hao sent me an interview request regarding an alleged misconduct case against Xiao-Qing Qiu of Sichuan University. According to Hao, Qiu had told her that the mass spectrometric analysis (MS) I did for his project verified his hypothesis that there was a "thiolactone ring" present in the protein pheromonicin. Hao asked me to explain to her in lay terms what I did and what the significance of this ring was. Hao's e-mail brought to my attention Qiu's paper, "An engineered multidomain bactericidal peptide as a model for targeted antibiotics against specific bacteria" (*1*). Reading the paper, I found that data from liquid chromatography–mass spectrometry (LC-MS) analysis were used to confirm the presence of the thiolactone ring in pheromonicin (p. 1481). I told Hao that I performed an MS analysis for Qiu at his request in 2003, but the results of the analysis I performed do not support the findings of the above-referenced article.

Qiu's stated interest with regard to the sample he provided to me in 2003 was, as above, in confirming the presence of the thiolactone ring in pheromonicin. On the basis of my memory and saved documents, his samples did not contain peptides at the predicted peptide masses within the mass measurement accuracy of the instrument or any masses matching the tryptic peptides of pheromonicin. I informed Qiu of this finding in early July of 2003. I do not know how Qiu obtained the MS data for his paper. However, I explained explicitly to Hao that the MS data presented in the paper have high mass measurement errors and should not have been used in the paper even if they were observed in mass spectra. The ultimate proof, of course, will be the reproducible production of the functional polypeptide based on Qiu's protocol.

HAITENG DENG

The Proteomics Resource Center, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA. E-mail: dengh@rockefeller.edu

Reference

1. X.-Q. Qiu, *Nat. Biotechnol.* **21**, 1480 (2003).

Extinction Risk and Conservation Priorities

THREATENED SPECIES LISTS BASED ON EXTINCtion risk are becoming increasingly influential for setting conservation priorities at regional, national, and local levels. Risk assessment, however, is a scientific endeavor, whereas priority setting is a societal process, and they should not be confounded (*1*). When establishing conservation priorities, it is important to consider financial, cultural, logistical, biological, ethical, and social factors in addition to extinction risk, to maximize the effectiveness of conservation actions.

The IUCN Red List Categories and Criteria (*2*) for assessing extinction risk are used through much of the world as an objective and systematic tool to develop regional, national, and local lists of threatened species (i.e., "Red Lists") [e.g., (*3*, *4*)]. Although it is widely recognized that a range of factors must be considered when establishing conservation priorities (*5*–*9*), a tendency still exists to assume that Red List categories represent a hierarchical list of priorities for conservation action and thus to establish conservation priorities based primarily, or even solely, on extinction risk. A survey of 47 national governments from around the world found that 82% of the countries that have or plan to prepare a national threatened species list are using these lists and/or the IUCN criteria in conservation planning and priority setting (*10*). Four of those countries automatically accord protected status to nationally threatened species. The actual number of countries that automatically and directly prioritize the most threatened species, without considering other factors, is undoubtedly greater.

Although extinction risk is a logical and essential component of any biodiversity conservation priority-setting system, it should not be the only one. While extinction risk assessment should be as objective as possible, priority setting must combine objective and subjective judgments, e.g. cultural preferences, cost of action, and likelihood of success (*4*, *8*, *9*). This process should not, however, be an excuse for lack of transparency. Effective priority-setting mechanisms should be explicit and include a rationale to justify the approaches taken.

REBECCA M. MILLER,[1] JON PAUL RODRÍGUEZ,[1,2]*
THERESA ANISKOWICZ-FOWLER,[3]
CHANNA BAMBARADENIYA,[4] RUBEN BOLES,[5]
MARK A. EATON,[6] ULF GÄRDENFORS,[7]
VERENA KELLER,[8] SANJAY MOLUR,[9] SALLY WALKER,[9]
CAROLINE POLLOCK[10]

[1]Centro de Ecología, Instituto Venezolano de Investigaciones Científicas, Apartado 21827, Caracas 1020-A, Venezuela. [2]Provita, Apartado 47552, Caracas 1041-A, Venezuela. [3]Species at Risk Branch, Canadian Wildlife Service, Environment Canada, Ottawa, ON K1A 0H3, Canada. [4]Asia Regional Species Programme, IUCN–The World Conservation Union, No. 53, Horton Place, Colombo 07, Sri Lanka. [5]COSEWIC Secretariat, c/o Canadian Wildlife Service, Ottawa, ON K1A 0H3, Canada. [6]The Royal Society for the Protection of Birds, The Lodge, Sandy, Bedfordshire, SG19 2DL, UK. [7]ArtDatabanken, Swedish Species Information Centre, Box 7007, S-750 07 Uppsala, Sweden. [8]Swiss Ornithological Institute, CH-6204 Sempach, Switzerland. [9]Zoo Outreach Organisation, 29-1 Bharathi Colony, First Cross, Peelamedu, PB 1683, Coimbatore, Tamil Nadu 641004, India. [10]IUCN/SSC Red List Programme, 219c Huntingdon Road, Cambridge, CB3 0DL, UK.

*To whom correspondence should be addressed. E-mail: jonpaul@ivic.ve

References

1. G. M. Mace, R. Lande, *Conserv. Biol.* **5**, 148 (1991).
2. IUCN, *Guidelines for Application of IUCN Red List Criteria at Regional Levels: Version 3.0* (IUCN Species Survival Commission, World Conservation Union, Gland, Switzerland, and Cambridge, UK, 2003).
3. F. Pinchera, L. Boitani, F. Corsi, *Biodivers. Conserv.* **6**, 959 (1997).
4. M. A. Eaton *et al.*, *Conserv. Biol.* **19**, 1557 (2005).
5. M. Avery *et al.*, *Ibis* **137**, S232 (1995).
6. V. Keller, K. Bollmann, *Conserv. Biol.* **18**, 1636 (2004).
7. U. Gärdenfors, *Trends Ecol. Evol.* **16**, 511 (2001).
8. R. D. Gregory *et al.*, *Br. Birds* **95**, 410 (2002).
9. J. P. Rodríguez, F. Rojas-Suárez, C. J. Sharpe, *Oryx* **38**, 373 (2004).
10. R. Miller *et al.*, *Report from the National Red List Advisory Group Workshop "Analysis of the Application of IUCN Red List Criteria at a National Level"* (World Conservation Union, Gland, Switzerland, 2005) (available at www.iucn.org/themes/ssc/red-lists.htm).

Confidentiality in Genome Research

THE POLICY FORUM ARTICLE "NO LONGER DEidentified" by A. L. McGuire and R. A. Gibbs (21 Apr., p. 370) discusses the importance of protecting privacy in genomic research and informing subjects of the privacy risks associated with public data-sharing in the consent process. In particular, the authors propose adopting a stratified consent process presenting three levels of confidentiality based on the number of single-nucleotide polymorphisms (SNPs) to be released.

It is necessary and crucial for all subjects to be fully informed about how their DNA data may be distributed, and to decide with whom they want their data shared. However, basing the decision to release data solely on the number of SNPs and their origin in single versus multiple gene loci is inadequate. The level of privacy risks posed by SNPs is also affected by many other factors, including linkage disequilibrium (LD) patterns among SNPs and frequencies of SNPs in the population.

Modest numbers of SNPs, especially those

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 6 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.

statistically independent ones, are as identifiable as social security numbers (*1*). Twenty statistically independent SNPs from single gene loci could pose more of a privacy threat than 75 SNPs with high LD from multiple gene loci. Even releasing eight SNPs can be risky for individuals with rare alleles, particularly if they are associated with a known phenotype. Therefore, it would be misleading to use arbitrary numbers of SNPs as a confidentiality indicator in the consent process. Nevertheless, we agree with the authors that sharing SNP data requires sufficient safeguards. Further risk assessment and strategy discussion will be needed.

ZHEN LIN,[1] RUSS B. ALTMAN,[2] ART B. OWEN[3]

[1]3 Smoketree Court, Durham, NC 27712–2690, USA. [2]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305–5120, USA. [3]Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA 94305–4065, USA.

Reference

1. Z. Lin, A. B. Owen, R. B. Altman, *Science* **305**, 183 (2004).

CFCs and the Size of the Ozone Hole

THE NETWATCH ITEM "OZONE TRACKER" (9 June, p. 1447) furthers the common misconception that the size of the Antarctic ozone hole is a function of ozone-destroying chlorofluorocarbons (CFCs). The column amount of ozone within the hole (its depth) may be controlled, in part, by inorganic chlorine derived from the breakup of CFCs, but the area occupied by the hole is not. Indeed, in the face of steadily rising amounts of atmospheric CFCs, the area has shrunk several times since 1979. It is cold wind-driven climatic conditions that create the polar vortex. This vortex isolates the atmosphere in the area of the hole, and polar stratospheric clouds forming within it may foster the deepening of the hole with destruction of the trapped ozone, but the total area covered by the vortex has nothing to do with CFCs.

KENNETH M. TOWE*

Department of Paleobiology, Smithsonian Institution, 230 West Adams Street, Tennille, GA 31089, USA.

*Senior Scientist Emeritus

CORRECTIONS AND CLARIFICATIONS

Letters: "Response" by Q. Lan *et al.* (19 May, p. 998). Because of an editing error, the reference list was numbered incorrectly. They are listed correctly here:

1. S. N. Yin *et al.*, *Br. J. Ind. Med.* **44**, 124 (1987).
2. N. Rothman *et al.*, *Cancer Res.* **57**, 2839 (1997).
3. Q. Lan *et al.*, *Cancer Res.* **65**, 9574 (2005).
4. T. Hastie *et al.*, *The Elements of Statistical Learning: Data Mining,Inference, and Prediction* (Springer-Verlag, Berlin, 2002).
5. H. Akaike, in *Second International Symposium on Information Theory*, B. N. Petrov, F. Csàki,Eds. (Akademia kiadó, Budapest, 1973), pp. 267–281.
6. S. Kim *et al.*, *Carcinogenesis*, 8 Dec. 2005; Epub ahead of print.

The reference numbers within the text are correct.

EVOLUTION

Dawkins's Dangerous Ideas

David C. Queller

Richard Dawkins has carved himself a very unusual niche in science. His books are intelligible and appealing to a popular audience but are also alive with ideas of interest to working scientists. The 30th anniversary of *The Selfish Gene* (*1*) is an apt occasion for *Richard Dawkins: How a Scientist Changed the Way We Think*, a celebratory volume in which Dawkins's students and colleagues line up to praise, extend, and occasionally contest his arguments. Fans of *The Selfish Gene* and Dawkins's other books can pick up and follow various strands of his legacy. The breadth of this legacy is reflected in the wide range of fields represented by the contributors: not just evolutionary biology and behavior, but psychology, computing, philosophy, religion (and skepticism), and even literature.

Richard Dawkins
How a Scientist Changed the Way We Think

Alan Grafen and Mark Ridley, Eds.

Oxford University Press, Oxford, 2006. 297 pp. $25, £12.99. ISBN 0-19-929116-0.

Among my personal favorites are two essays that together bookend Dawkins's talents. At one end, the novelist Philip Pullman celebrates Dawkins's writing: the personal touch, the narrative drive, the memorable phrases—in short, the "gift for combining words in a knot that stays tied." At the other end there is Alan Grafen's exposition of the intellectual merit of *The Selfish Gene*. It was not just a confection of memorable phrases but fresh thinking on how the concepts of replicators and selfishness bring together the new theories on social evolution. Like Darwin, Dawkins worked in nonmathematical terms. Unlike Darwin, he had to contend with a skeptical mathematical priesthood. He succeeded because he also had a gift for using logic in a way that stays tied.

Andrew Read opens the volume with an account of how his view of life was changed after reading *The Selfish Gene* on a lonely mountaintop in New Zealand. My own first reading had less of Mt. Sinai in it but was still special. I was in the flats of Michigan in my first year of grad school, and Richard Alexander and John Maynard Smith were already laying waste to the false idol of uncritical group selection. Alerted by Maynard Smith to the imminent appearance of *The Selfish Gene*, I watched for it, snapped it up immediately, and, though I am neither a night owl nor a rapid reader, I had devoured it whole by the early hours of the next morning. Although I was familiar with many of the ideas, Dawkins crystallized the logic of the new theories and pushed them deeper with his unrelenting gene-centered approach.

My enthusiasm was not universally shared. I persuaded my father, a historian of medieval Venice, to read the book. To my dismay, he pronounced it—I think this was his word—obscene. I suspect he was partly repulsed by the metaphors (for example, that we are all lumbering robots). Despite Dawkins's repeated cautions, readers tended to take these too literally. But even without the vivid metaphors, the message is disturbing enough. Here was Darwin's materialism applied to that which we hold most dear: how we treat, and are treated by, our neighbors, friends, and families. And here Dawkins offers the only thing worse than Darwin's purposeless universe: a universe driven by the seemingly malevolent egoism of hereditary molecules. Genes could not be altruistic; any sacrifice must be repaid by a greater fitness benefit, or by benefits to kin who have copies of the gene.

Prophet of the selfish gene.

CREDIT: COLIN MCPHERSON/CORBIS

In his chapter, the philosopher Daniel Dennett recalls how, hearing unfavorable comments about the book, he missed out on reading *The Selfish Gene* for several years. I am sure he and most of the contributors would advise readers not to put off reading the real thing even in favor of their own admiring chapters. Indeed, several contributors note that *The Selfish Gene* bears rereading even after all these years. I just reread both *The Selfish Gene*, perhaps my favorite nonfiction book of my college years, and its counterpart on my fiction list, *Catch-22* (*2*), and there were some curious resonances. In *Catch 22*, Yossarian's plight is a classic social dilemma. As a bombardier in World War II, he believed in the justice of the Allied cause. But he also believed that the Allies would win whether he continued to fly dangerous missions or not, and he preferred not to be among the dead. When asked "But what if everyone thought that way?" he would reply "Then I'd be crazy to think anything else, wouldn't I?" Yossarian, perhaps following the dictates of his selfish genes, did not want the sucker's payoff.

Each book revolves around a dark secret. In the novel, Yossarian's motivation is gradually revealed in the story of his mission over Avignon. The tail gunner, Snowden, has been hit, and Yossarian is relieved to be able to neatly dress the flak wound in his leg. But then Snowden spills his dirty secret from beneath his flak jacket, in the form of a second wound—gaping, twitching, hopelessly mortal. The secret he forced upon Yossarian, no less powerful for being known in advance, is that all humanity is flesh—fragile, mortal flesh.

Dawkins spills his own dirty, obscene secret, again no less powerful now that we have known it for 30 years. All flesh is survival machinery, and the survival it promotes is that of our selfish genes. In the volume under review, the psychiatrist Randolph Nesse gives a kind of talking cure for those traumatized by Dawkins's secret, but he admits that it may not suffice. If humanity has struggled since at least the Neandertals with Yossarian's dirty secret of mortality, then we may take a while to adjust to the one that Dawkins spilled.

But there is a difference. Yossarian's secret is the fact of mortality, whereas Dawkins's secret is a theory. It is not the difference in levels of certainty that is crucial, for I am confident that Dawkins's theory is essentially correct. It is instead that the facts of sociality, including human sociality, are prior to any theory. We already knew that humans display a baffling mixture of good and evil, of cooperation and egoism. For example, nothing is more evil than war, but that is made possible only by extreme cooperation and sacrifice by selfless non-Yossarians. The facts of the social world are not changed by Dawkins. Rather, as the book's subtitle says, he changed the way we think about it and provided us with tools to try to understand it. In my rereading of *The Selfish Gene*, I found that a bit of the original frisson

The reviewer is at the Department of Ecology and Evolutionary Biology, Rice University, Post Office Box 1892, Houston, TX 77251–1892, USA. E-mail: queller@rice.edu

had faded and that what remained were good, sensible ways to try to comprehend our world. I think even my father came to agree, at least in part; before he died he had set to work studying the importance of kinship and nepotism in his medieval Venetians.

References

1. R. Dawkins, *The Selfish Gene* (Oxford Univ. Press, Oxford, 1976).
2. J. Heller, *Catch-22* (Simon and Schuster, New York, 1961).

10.1126/science.1130993

ENGINEERING

Shaking and Shaping San Francisco

Christine Theodoropoulos

Bracing for Disaster
Earthquake-Resistant Architecture and Engineering in San Francisco, 1838–1933

by Stephen Tobriner

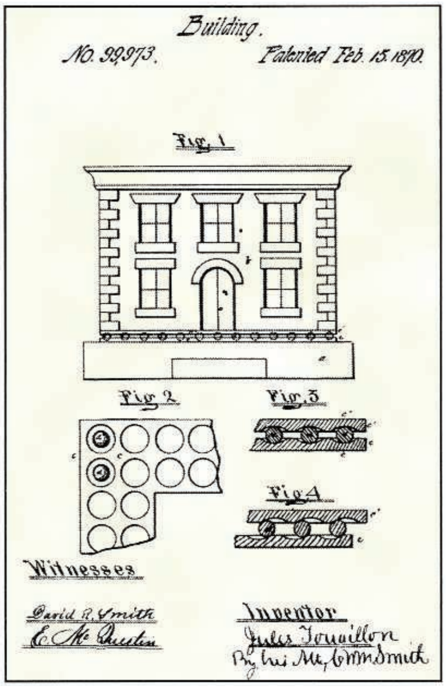
Heyday and Bancroft Library, University of California, Berkeley, CA, 2006. 351 pp. Paper, $30. ISBN 1-59714-025-2.

In the aftermath of urban earthquakes, how do architects and engineers use the lessons they learn to rebuild safer cities? How do citizens, and the financial and governmental entities responsible for reconstruction, support design and construction practices that produce better performance in future earthquakes? As we commemorate the centennial anniversary of the 1906 San Francisco earthquake, it is important to recognize that natural disasters are among the processes that shape cities. With a full century of hindsight, it is also time to reconsider past interpretations of the history of earthquake-resistant building practices.

Although the events of 18 April 1906 did much to raise awareness of the risks of building in earthquake country, efforts to rebuild the devastated city have often been cited as negative examples that ignored the seismic threats to San Francisco. The fires masked evidence of earthquake-induced damage. Social and economic pressures promoted quick rebuilding. San Francisco's building codes were not revised to include new seismic provisions, and the use of unreinforced brick masonry continued. Thus, many analysts have concluded that the need for rapid recovery using existing technology within the limitations of engineering knowledge perpetuated building practices that caused the city to rebuild in a manner that disregarded earthquake-resistant design. Most American histories of earthquake engineering begin later in the 20th century, when the earliest seismic code provisions were written in response to the 1925 Santa Barbara earthquake and building damage observed during the 1933 Long Beach earthquake led California to mandate the first statewide regulations.

But, as Stephen Tobriner argues in *Bracing for Disaster*, a closer look at building design and construction practices in late 19th- and early 20th-century San Francisco reveals efforts to build urban structures suited to earthquake country. During these decades, the challenges of seismic design were actively addressed as architects, engineers, and builders responded to the desire of owners, insurers, and government to reduce earthquake risks. Tobriner (an architectural historian at the University of California, Berkeley, and San Francisco native) presents evidence gleaned from historic photographs, construction documents, and observations of buildings (including hidden details revealed during demolitions) as well as searches through archives that portray civic and professional dialogues concerning the earthquake problem. This documentation, combined with a careful rereading of the construction history of San Francisco, indicates that earthquake engineering practice in the United States began earlier and incorporated greater insight into building performance than reported in prior histories.

Tobriner's fascinating account of several innovative "earthquake-proof" construction systems introduced after the 1865 and 1868 San Francisco earthquakes reveals that 19th-century inventors had begun to recognize many of the seismic design principles that form the basis of today's engineering practice. Patented schemes for incorporating horizontal bands and vertical bars of bond iron into masonry walls and a system of external iron bracing for masonry houses are precedents for later reinforced masonry technology. Although the use of base isolation technology is a relatively recent development (dating from the 1990s), in 1870 Jules Touaillon (an otherwise unknown San Francisco resident and inventor) was awarded a patent for a base isolator constructed of load-bearing balls free to roll within indentations in plates placed between a building and its foundation. The revolutionary idea of accommodating, rather than resisting, movement in a building structure is revealed in another example of innovative engineering: in his design for the politically charged 1912 City Hall project, Christopher Snyder included a shock-absorbing flexible first story (which has been credited with saving the building from collapse during the 1989 Loma Prieta earthquake). Throughout the book, Tobriner uses building case studies to place the use of earthquake-resistant technologies in context and explain the connections between engineering design decisions, architectural design objectives, and the perspectives of stakeholders.



Avant-garde solution to shaking. In his U.S. patent (1870) for base isolation, Jules Touaillon proposed building brick structures on platforms that rest on balls, each of which can roll within a constrained space.

CREDIT: FROM *BRACING FOR DISASTER*

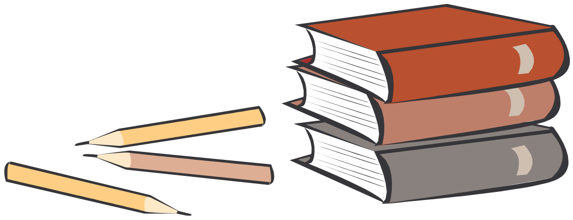
Although the dramatically visible damage to building structures generally receives the majority of attention in the aftermath of urban earthquakes, cities are more than collections of buildings. Urban form responds to the natural systems of topography, soils, and water. It is shaped by the way nature interacts with urban infrastructures that support the quality of urban life and protect public health and safety. Tobriner's account of the history of San Francisco's earthquakes examines connections between earthquake experience and urban form. His discussions of the reshaping of topography to accommodate transportation and growth, responses to the threat of urban fires, economic impacts of insurance company practices, and the development of water supply systems provide readers with an understanding of the interaction between earthquakes and urban systems. Extensively illustrated with annotated photographs, maps, and drawings that invite the reader to interpret physical evidence, *Bracing for Disaster* presents a unique history of a unique city. Add a map of today's San Francisco, and the book also functions as an informative guidebook to the city as seen through the lens of earthquake-resistant design.

10.1126/science.1130008

The reviewer is at the Department of Architecture, University of Oregon, Eugene, OR 97403–1206, USA. E-mail: ctheodor@uoregon.edu

PROFESSIONAL DEVELOPMENT

Who Is Responsible for Preparing Science Teachers?



Valerie Otero,[1]* Noah Finkelstein,[2] Richard McCray,[3] Steven Pollock[2]

At the University of Colorado at Boulder, involving students in the transformation of science courses raises the visibility of science teaching as a career and produces K–12 teachers well-versed in science.

Teachers knowledgeable in both science and pedagogy are critical for successful math and science education in primary and secondary schools. However, at U.S. universities, too many undergraduates are not learning the science (*1*–*3*), and our highest performing students are choosing fields other than teaching (*4*). With a few exceptions [such as (*5*, *6*)], universities convey that teaching kindergarten to 12th grade (K–12) is not a career worthy of a talented student (*7*). Two out of three high school physics teachers have neither a major nor a minor in the discipline (*8*), and the greatest teacher shortages are in math, physics, and chemistry. The shortages of teachers with these majors have likely contributed to the poor current outcomes (*9*) for math and science education [supporting online material (SOM) text].

Enhanced online at www.sciencemag.org/cgi/content/full/313/5786/445

The first of four recommendations by the National Academies for ensuring American competitiveness in the 21st century was to "increase America's talent pool by vastly improving K–12 science and mathematics education" (*9*). Teacher preparation is not solely the responsibility of schools of education. Content knowledge is one of the main factors positively correlated with teacher quality (*10*), yet the science faculty members directly responsible for teaching undergraduate science are rarely involved in teacher recruitment and preparation.

The Learning Assistant Model

At the University of Colorado (CU) at Boulder, we have developed a program that engages both science and education faculty in addressing national challenges in education. Undergraduate learning assistants are hired to assist science faculty in making their courses student centered, interactive, and collaborative—factors that have been shown to improve student performance (*1*–*3*). The program also recruits these learning assistants to become K–12 teachers. Thus, efforts to improve undergraduate education are integrated with efforts to recruit and prepare future K–12 science teachers.

Since the program began in 2003, we have transformed 21 courses (table S1) with the participation of 28 science and math faculty members, 4 education faculty members, and 125 learning assistants. The learning assistants support and sustain course transformation—characterized by actively engaged learning processes—by facilitating collaboration in the large-enrollment science courses (fig. S1). The program also increases the teacher-to-student ratio by a factor of 2 to 3 (SOM text). Without learning assistant participation, such courses tend to be dominated by the lecture format. Faculty members new to course transformation are supported by faculty that have experience working with learning assistants (SOM text).

About 50 learning assistants have been hired each semester for courses in six departments: physics; astrophysical and planetary sciences; molecular, cellular, and developmental biology (MCD biology); applied mathematics; chemistry; and geological sciences. The learning assistants are selected through an application and interview process according to three criteria: (i) high performance as students in the course; (ii) interpersonal and leadership skills; and (iii) evidence of interest in teaching. Learning assistants participate as early as the second semester of freshman year and as late as senior year. Learning assistants differ from traditional teaching assistants (TAs) in that learning assistants receive preparation and support for facilitating collaborative learning.

Learning assistants receive a modest stipend for working 10 hours per week in three aspects of course transformation. First, learning assistants lead learning teams of 4 to 20 students that meet at least once per week. Learning assistant–led learning teams work on collaborative activities ranging from group problem-solving with real astronomical data to inquiry-based physics activities. Second, learning assistants meet weekly with the faculty instructor to plan for the upcoming week, to reflect on the previous week, and to provide feedback on the transformation process. Finally, learning assistants are required to take a course on Mathematics and Science Education that complements their teaching experiences. In this course, cotaught by a faculty member from the School of Education and a K–12 teacher, learning assistants reflect on their own teaching, evaluate the transformations of courses, and investigate practical techniques and learning theory (SOM text).

Through the collective experiences of teaching as a learning assistant, instructional planning with a science faculty member, and working with education faculty, learning assistants develop pedagogical content knowledge, which is characteristic of effective teachers (*11*). The skills that learning assistants develop are valuable for teaching at all levels and in many environments. Those learning assistants who consider K–12 teaching as a career are encouraged to continue and are eligible for NSF-funded Noyce Teaching Fellowships (fig. S2).

Undergraduates enrolled in science teacher certification programs

Major	All of Colorado (2004–2005) LAs not included	CU Boulder (2004–2005) LAs not included	CU Boulder (2005–2006) LAs recruited
Physics and astrophysics	2	1	7
MCD biology	0	0	4
Chemistry	14	0	N.A.
Geoscience	11	0	N.A.

More students enticed into teaching. The learning assistant (LA) program at CU Boulder improved recruitment of undergraduate students into K–12 teacher certification programs relative to the undergraduate recruitment rates noted for 2004 to 2005 without the learning assistant program. Chemistry and geoscience joined the program in 2006, and so have not yet recruited students into teaching certification programs. N.A., not applicable.

Results of the Learning Assistant Program

The learning assistant program has successfully increased the number and quality of future science teachers, improved student understanding of science content, and engaged a broad range of science faculty in course transformation and teacher education.

To date, 125 math and science majors have participated as learning assistants and 18 of

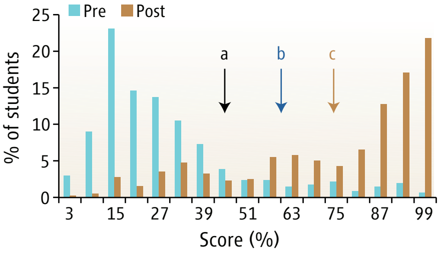
[1]School of Education, [2]Department of Physics, [3]Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, CO 80309, USA.

*To whom correspondence should be addressed. E-mail: valerie.otero@colorado.edu

Published by AAAS

them (6 math and 12 science) have joined teacher certification programs. These learning assistants have an average cumulative grade point average (GPA) of 3.4, higher than the typical 2.9 GPA for math and science majors who express interest in teaching (*12*). In physics at CU Boulder, the average GPA for majors is 3.0, and it is 3.75 for learning assistants.

The learning assistant program improved recruitment rates to science teacher certification programs over preexisting rates (see table on page 445). Before the learning assistant program, about two students per year from our targeted science majors enrolled in certification programs. Nationwide, about 300 physics majors each year are certified to teach (*13*). Thus, even small improvements in recruitment rates could have an impact on the pool of available teachers, particularly in the state of Colorado (*14*). Most of the learning assistants who decided to become teachers report that they had not explored teaching as a career until participating as learning assistants. Factors that led to decisions to become teachers include recognition of teaching as intellectually challenging and positive attitudes among participating faculty (*7*).



Learning assistants improve student learning. Pretest and posttest FMCE results for CU students in a transformed course with learning assistants. The pretest median is 24% (±1%) (n = 467); the posttest median is 85% (±1%) (n = 399). Arrows indicate posttest average (mean) scores for (a) students nationwide in traditional courses with pretest scores matching those of CU students, (b) students in a CU course that features educational reforms but no learning assistants, and (c) students in the CU course transformed with learning assistants (arrow shows the mean of the brown bars).

Development of Content Knowledge

Each of the participating departments demonstrates improved student achievement as a result of the learning assistant program (*15*–*17*). The transformation of the introductory calculus-based physics sequence provides an example. These courses are large (500 to 600 students), with three lectures per week implementing peer instruction and personal response systems (*17*, *18*). The learning assistant program has provided enough staff to implement student-centered tutorials with small-group activities (*19*). Learning assistants and TAs train together weekly to circulate among student groups and ask guiding questions. The number of applicants for learning assistant positions in physics is currently 50 to 60 per term for 15 to 20 positions.

We assessed student learning with the Force and Motion Concept Evaluation (FMCE) (*20*) and the Brief Electricity and Magnetism Assessment (BEMA) (*21*). In transformed courses, students had an average normalized improvement of 66% (±2% SEM) for the FMCE test (see chart, left), nearly triple national average gains found for traditional courses (*3*, *22*). With the BEMA exam, the average normalized learning gains for students in the transformed courses ranged from 33 to 45%. National averages are not yet available for this new BEMA exam. The normalized learning gains for the learning assistants themselves average just below 50%, with their average posttest score exceeding average scores for incoming physics graduate students. In a different model, students enrolled in a physics education course can opt to participate as learning assistants for additional credit (*23*). These students make gains twice that of their peers who do not opt to participate as learning assistants. Students who engage in teaching also demonstrate increased understanding of the nature of teaching and improved abilities to reflect on their understanding of teaching and learning (*23*) (table S2).

Impact on Faculty

Faculty members participating in the learning assistant program have started to focus on educational issues not previously considered. Faculty members report increased attention to what and how students learn. In a study of faculty response to this program, all 11 faculty members interviewed reported that collaborative work is essential, and learning assistants are instrumental to change (*7*). One faculty member notes: "I've taught [this course] a million times. I could do it in my sleep without preparing a lesson. But [now] I'm spending a lot of time preparing lessons for [students], trying to think 'Okay, first of all, what is the main concept that I'm trying to get across here? What is it I want them to go away knowing?' Which I have to admit, I haven't spent a lot of time in the past thinking about." This type of statement is common among those who engage in course transformation for the first time (SOM text).

Sustaining Successful Programs

The learning assistant model can be sustained and modified for a variety of institutional environments. Another longstanding successful model, the UTeach program at the University of Texas (*5*) has demonstrated that it is possible to internally sustain educational programs for science majors. These and other model programs bring together partners who each have a vested interest in increasing the number of high-quality teachers and the number of math and science majors, as well as improving undergraduate courses.

Implementation of a learning assistant program requires local interest from faculty in the sciences and education, as well as administrative backing and funding of a few thousand dollars per learning assistant per year (SOM text). The cost of a learning assistant is less than one-fifth that of a graduate TA. Learning assistants may also receive credit in lieu of pay. Another model is to fund learning assistant stipends from student fees.

With collective commitment, education can be brought to greater visibility and status, both for students considering teaching careers and for faculty teaching these students (SOM text). As scientists, we can address the critical shortfall of K–12 science teachers by improving our undergraduate programs and supporting interest in education.

References and Notes

1. J. Handelsman *et al.*, *Science* **304**, 521 (2004).
2. J. Handelsman *et al.*, *Science* **306**, 229 (2004).
3. R. Hake, *Am. J. Phys*. **66**, 64 (1998).
4. National Science Board, *Science and Engineering Indicators 2006* [National Science Foundation (NSF), Arlington, VA, 2006], vol. 1, NSB 06-01; vol. 2, NSB 06-01A.
5. UTeach (https://uteach.utexas.edu).
6. Physics Teacher Education Coalition (www.ptec.org).
7. V. Otero, paper presented at the AAAS Annual Meeting, Washington, DC, 17 to 21 February 2005.
8. M. Neuschatz, M. McFarling, *Broadening the Base: High School Physics Education at the Turn of the New Century* (American Institute of Physics, College Park, MD, 2003).
9. NRC, *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Future* (National Research Council, Washington, DC, 2005).
10. U.S. Department of Education, Office of Policy Planning and Innovation, *Meeting the Highly Qualified Teachers Challenge: The Secretary's Second Annual Report on Teacher Quality* (Editorial Publications Center, Washington, DC, 2002).
11. L. S. Shulman, *Educ. Res*. **15**, 4 (1986).
12. L. Moin *et al.*, *Sci. Educ*. **89**, 980 (2005).
13. M. Neuschatz, personal communication.
14. Colorado Commission of Higher Education, *Report to Governor and General Assembly on Teacher Education* (CCHE, Denver, CO, 2006).
15. J. K. Knight, W. B. Wood, *Cell Bio. Educ*. **4**, 298 (2005).
16. M. Nelson, doctoral dissertation, University of Colorado, Boulder, CO (2005).
17. N. D. Finkelstein, S. J. Pollock, *Phys. Rev. ST Phys. Educ. Res*. 1, 010101 (2005).
18. E. Mazur, *Peer Instruction: A User's Manual* (Prentice-Hall, Englewood Cliffs, NJ, 1997).
19. L. McDermott, P. Shaffer, *Physics Education Group, Tutorials in Introductory Physics* (Prentice-Hall, Saddle River, NJ, 2002).
20. R. K. Thornton, D. R. Sokoloff, *Am. J. Phys*. **66**, 338 (1998).
21. L. Ding, R. Chabay, B. Sherwood, R. Beichner, *Phys. Rev. ST Phys. Educ. Res*. **2**, 010105 (2006).
22. The student normalized improvement is defined as (posttest – pretest)/(100 – pretest).
23. N. D. Finkelstein, *J. Sch. Teach. Learn*. **4**, 1 (2004).
24. This work is supported by the NSF, the American Institute of Physics, the American Physical Society, the American Association of Physics Teachers, and the University of Colorado. We thank the STEM Colorado team and the PER group at the CU Boulder for helping develop and maintain this effort.

10.1126/science.1129648

Supporting Online Material
www.sciencemag.org/cgi/content/full/313/5786/445/DC1

MOLECULAR BIOLOGY

Self-Correcting Messages

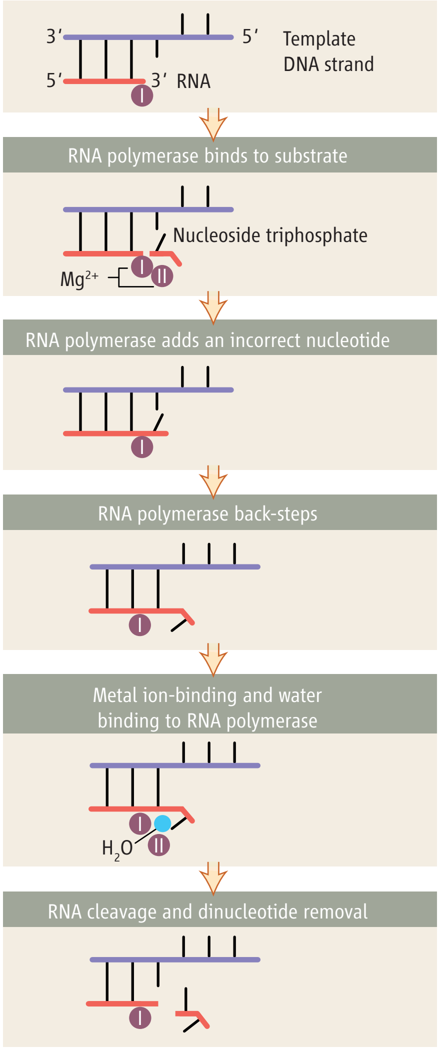
Patrick Cramer

Mistakes can occur as RNA polymerase copies DNA into transcripts. A proofreading mechanism that removes the incorrect RNA is triggered by the erroneous RNA itself.

Precision can be vital. Living cells transcribe their DNA genomes into messenger RNA (mRNA), which then directs protein synthesis. These processes are not without mistakes, but cells have evolved processes for proofreading and correction to shut down the propagation of errors. On page 518 of this issue, Zenkin *et al.* report that mRNA itself helps correct errors that occur during its own synthesis (*1*). This finding helps to explain the fidelity of gene transcription and suggests that self-correcting RNA was the genetic material during early evolution.

During gene transcription, the enzyme RNA polymerase moves along the DNA template and synthesizes a complementary chain of ribonucleotides, the mRNA. Errors arise when the growing mRNA incorporates a nucleotide that is not complementary to the DNA template. Nucleotides could, in principle, be removed by an RNA cleavage activity of the polymerase (*2*), but this intrinsic activity is very weak. Transcript cleavage factors enhance the polymerase's cleavage activity, and render error correction efficient in vitro (*3*, *4*). These cleavage factors are, however, not essential in vivo. These observations have led to the widespread belief that transcriptional error correction may not be critical for cellular function. However, erroneous mRNA could produce nonfunctional or harmful proteins, arguing for the existence of a mechanism that increases transcriptional fidelity.

Zenkin *et al.* now describe a simple mechanism for efficient, factor-independent error correction during transcription (see the figure). The authors assembled complexes of bacterial RNA polymerase with synthetic DNA and RNA. The RNA chains contained at their growing end either a nucleotide complementary to the DNA template, or a noncomplementary nucleotide that mimicked the result of misincorporation. In a key experiment, addition of magnesium ions triggered efficient cleavage from a polymerase-DNA-RNA complex of an RNA dinucleotide containing an erroneous nucleotide, but not from error-free complexes. Further biochemical experiments showed that RNA polymerase within an erroneous complex slides backwards or "back-steps" along DNA and RNA, and that the terminal, noncomplementary nucleotide participates in catalyzing removal of itself, together with the penultimate nucleotide. When the experiments were repeated in the presence of nucleoside triphosphates, the substrates for RNA synthesis, most of the RNA in erroneous complexes was still cleaved, although a fraction of the RNA was extended past the misincorporation site. Thus, RNA-stimulated RNA cleavage after misincorporation may suffice for transcriptional proofreading.


3′
5′
Template
DNA strand
5′
3′ RNA
I
RNA polymerase binds to substrate
Nucleoside triphosphate
Mg^{2+}
I
II
RNA polymerase adds an incorrect nucleotide
I
RNA polymerase back-steps
I
Metal ion-binding and water
binding to RNA polymerase
I
H_2O
II
RNA cleavage and dinucleotide removal
I

RNA-assisted transcriptional proofreading. Correction of misincorporation errors at the growing end of the transcribed RNA is stimulated by the misincorporated nucleotide. Mg^{2+} ions are bound to the catalytic region of RNA polymerase.

What is the chemical basis for such observed transcriptional proofreading? Both RNA synthesis and RNA cleavage occur at a single, highly conserved active site (*5*–*8*), and require two catalytic magnesium ions (*5*, *9*–*12*). The first metal ion is persistently bound in the active site, whereas the second is exchangeable. Binding of the second metal ion is stabilized by a nucleoside triphosphate during RNA synthesis, or by a transcript cleavage factor during RNA cleavage. Zenkin *et al.* show that the base of the back-stepped misincorporated nucleotide can also stabilize binding of the second metal ion (*1*). In addition, the misincorporated nucleotide and transcript cleavage factors may both activate a water molecule that acts as a nucleophile in the RNA cleavage reaction. Thus, the terminal RNA nucleotide plays an active role in RNA cleavage.

These results strengthen and extend the model of a multifunctional, "tunable" active site in RNA polymerases. Nucleoside triphosphates, cleavage factors, and back-stepped RNA can occupy similar locations in the active site, and position the second catalytic metal ion for RNA synthesis or cleavage. Because RNA dinucleotides are generally obtained in the presence of cleavage factors, the terminal RNA nucleotide and a cleavage factor likely cooperate during RNA cleavage from a back-stepped state. If the RNA is further backtracked, cleavage factors become essential for RNA cleavage, because the terminal nucleotide is no longer in a position to stimulate cleavage. In both scenarios, RNA cleavage provides a new, reactive RNA end and a free adjacent substrate site, allowing transcription to resume.

The discovery of self-correcting RNA transcripts suggests a previously missing link in molecular evolution (*13*). One prerequisite of an early RNA world (devoid of DNA) is that RNA-based genomes were stable. Genome stability required a mechanism for RNA replication and error correction during replication, which could have been similar to the newly described RNA proofreading mechanism described by Zenkin *et al.* If self-correcting replicating RNAs coexisted with an RNA-based protein synthesis activity, then an early RNA-based replicase could have been replaced by a protein-based RNA replicase. This ancient protein-based RNA replicase could have evolved to accept DNA as a template, instead of RNA, allowing the transition from RNA to DNA genomes. In this scenario, the resulting DNA-dependent RNA polymerase retained the ancient RNA-based RNA proofreading mechanism.

Whereas an understanding of RNA proof-

The author is at the Gene Center Munich, Department of Chemistry and Biochemistry, Ludwig-Maximilians-Universität München, Feodor-Lynen-Strasse 25, 81377 Munich, Germany. E-mail: cramer@lmb.uni-muenchen.de

Published by AAAS

reading is only now emerging, DNA proofreading had long been characterized. DNA polymerases cleave misincorporated nucleotides from the growing DNA chain, but the cleavage activity resides in a protein domain distinct from the domain for synthesis (*14*). The spatial separation of the two activities probably allowed optimization of two dedicated active sites during evolution, whereas RNA polymerase retained a single tunable active site. This could explain how some DNA polymerases achieve very high fidelity, which is required for efficient error correction during replication of large DNA genomes.

In the future, structural studies will unravel the stereochemical basis for RNA proofreading. Further biochemical and single-molecule studies should clarify how back-stepping and other rearrangements at the tunable polymerase active site are triggered. Techniques must also be developed to probe the in vivo significance of different aspects of the transcription mechanism discovered in vitro.

References

1. N. Zenkin, Y. Yuzenkova, K. Severinov, *Science* **313**, 518 (2006).
2. M. Orlova, J. Newlands, A. Das, A. Goldfarb, S. Borukhov, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4596 (1995).
3. M. J. Thomas, A. A. Platas, D. K. Hawley, *Cell* **93**, 627 (1998).
4. D. A. Erie, O. Hajiseyedjavadi, M. C. Young, P. H. von Hippel, *Science* **262**, 867 (1993).
5. V. Sosunov *et al.*, *EMBO J.* **22**, 2234 (2003).
6. H. Kettenberger, K.-J. Armache, P. Cramer, *Cell* **114**, 347 (2003).
7. N. Opalka *et al.*, *Cell* **114**, 335 (2003).
8. V. Sosunov *et al.*, *Nucleic Acids Res.* **33**, 4202 (2005).
9. P. Cramer, D. A. Bushnell, R. D. Kornberg, *Science* **292**, 1863 (2001).
10. T. A. Steitz, *Nature* **391**, 231 (1998).
11. D. G. Vassylyev *et al.*, *Nature* **417**, 712 (2002).
12. K. D. Westover, D. A. Bushnell, R. D. Kornberg, *Cell* **119**, 481 (2004).
13. A. M. Poole, D. T. Logan, *Mol. Biol. Evol.* **22**, 1444 (2005).
14. L. S. Beese, T. A. Steitz, *EMBO J.* **10**, 25 (1991).

10.1126/science.1131205

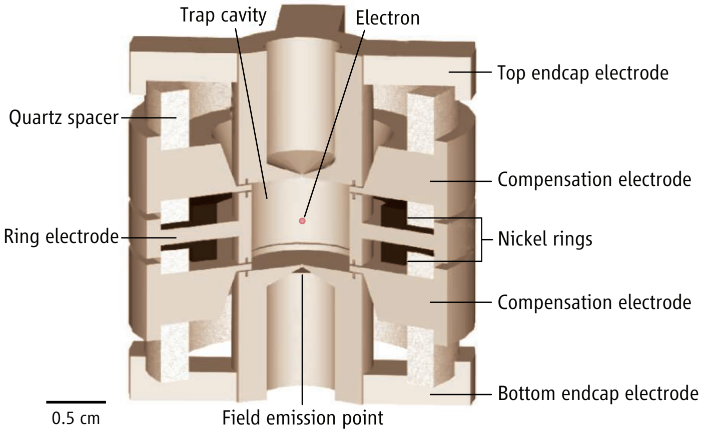
PHYSICS

A More Precise Fine Structure Constant

The fine structure constant, a vital quantity in quantum theory, sets the scale for the physical world. Recent measurements have improved its precision by a factor of 10.

Daniel Kleppner

Relativistic quantum electrodynamics (QED)—the theory that describes electromagnetic interactions between all electrically charged particles—is the most precisely tested theory in physics. In studies of the magnetic moment of the electron (a measure of its intrinsic magnetic strength), theory and experiment have been shown to agree within an uncertainty of only 4 parts per trillion. This astounding precision has just been improved. A new measurement by Odom *et al*. (*1*) has increased the experimental precision by a factor close to 6. In a parallel theoretical effort, Gabrielse *et al.* (*2*) have extended the QED calculations of the magnetic moment to a new level of precision. By combining these advances, the precision with which we know the value of the fine structure constant is now 10 times as high as that obtained by any other method. The fine structure constant is a dimensionless number, ~1/137, which involves the charge of the electron, the speed of light, and Planck's constant. It is usually designated α, and it plays a ubiquitous role in quantum theory, setting the scale for much of the physical world. Thus, α occupies an honored position among the fundamental constants of physics.



One-electron cyclotron. A magnetic field along the axis confines the electron radially; an oscillating electric field applied to the endcap electrodes confines it longitudinally. Nickel rings slightly perturb the magnetic field so as to couple the radial and longitudinal motions. The electron is trapped in a cavity that inhibits spontaneous emission. Other electrodes are used to control the electric field so as to reduce QED effects of the vacuum.

CREDIT: G. BABRIELSE/HARVARD UNIVERSITY

The quantity that has been measured by these researchers is the ratio of the magnetic moment of the electron to the fundamental atomic unit of magnetism known as the Bohr magneton. This dimensionless ratio is called the *g*-factor of the electron. Because the *g*-factor is a basic property of the simplest of the elementary particles, it has played a prominent role both in motivating and testing QED. According to Dirac's theory of the electron (*3*, *4*), for which he received the Nobel Prize in 1933, the *g*-factor should be exactly 2. In the period immediately following World War II, new data on the spectrum of hydrogen led to the creation of QED by Schwinger, Feynman, Tomonaga, and Dyson (*5*). According to QED, the electron *g*-factor would differ slightly from 2. Kusch and Foley discovered experimentally that the *g*-factor differed from 2 by about 1 part in a thousand (*6*). For this work Kusch received the Nobel Prize in 1955, followed by Schwinger, Feynman, and Tomonaga, who received the Nobel Prize in 1965. In 1987 Dehmelt published the measurement referred to above, accurate to 4 parts per trillion, for which he received the Nobel Prize in 1989 (*7*). The major experimental innovation in Dehmelt's measurement was a technique that allowed him to observe a single electron. The experiment of Gabrielse and colleagues builds on Dehmelt's work but incorporates major innovations that make the isolated electron into a quantum system whose energy levels can be probed.

The experiment compares the two types of motion of an electron in a magnetic field. The first is circular motion around the direction of the field at a frequency known as the cyclotron frequency f_c because the motion is described by the same equation as that for charged particles in a cyclotron accelerator. The second type of motion is spin precession. An electron possesses intrinsic spin, somewhat in analogy to the spin of a flywheel in a gyroscope. If a gyroscope is suspended by one end of its axle, it

The author is in the Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. E-mail: kleppner@mit.edu

Published by AAAS

experiences a torque due to its weight and precesses about a vertical axis. Similarly, in a magnetic field, an electron experiences a torque due to its magnetic moment, and the electron spin axis precesses about the field at a frequency f_s. The *g*-factor differs from 2 by the ratio $(f_s - f_c)/f_c$. The quantities actually measured are the cyclotron frequency f_c and the difference frequency $(f_s - f_c)$.

To carry out the measurement, Gabrielse and co-workers designed a one-electron cyclotron in which the underlying quantum nature of the electron's motion is both exploited and controlled (see the figure). In the theory of QED, the vacuum plays an important dynamical role. The radiation field of the vacuum (a fluctuating field in totally empty space) is a principal source of the electron moment anomaly. The vacuum field is slightly affected by conducting surfaces, such as the electrodes in the one-electron cyclotron. By carefully controlling the geometry of the cyclotron, Gabrielse and his colleagues essentially eliminated perturbation of the *g*-factor by the vacuum. Using principles of cavity QED, the researchers arranged the geometry so as to substantially prevent the orbiting electron from radiating its energy, thereby lengthening the observation time of each measurement.

Because cyclotron motion is inherently quantized, the energy of a circulating charged particle can change only in steps of hf_c, where *h* is Planck's constant. Normally these energy steps are so small compared to the particle's energy that the underlying quantum nature of the motion is unimportant. In the quantum one-electron cyclotron, however, the energy is so finely controlled that each discrete step can be observed. To accomplish this, the research team had to eliminate effects of thermal radiation by carrying out the experiment at a temperature of 0.1 K. Under these conditions, and using a technique called quantum jump spectroscopy, they could clearly see whether the electron was in the ground cyclotron energy state, or had taken one, two, or more energy steps.

An intriguing feature of the one-electron cyclotron is that the energy steps are not exactly equal due to the relativistic shift of the electron's mass with energy. One would hardly expect relativity to play a role at the ultralow energy of the one-electron cyclotron, but at the scale of precision of the experiment, relativistic effects are important. Odom *et al*. measured $g/2 = 1.00115965218085$, with an uncertainty of only 7.6 parts in 10^{13}, or 0.76 parts per trillion (*1*).

Calculation of the electron moment anomaly with the theory of QED presents a formidable challenge. The calculation involves evaluating the coefficients of terms in a power series, with each new term much more complex than the previous one. The third-order term was calculated in the mid-1990s (*8*). The fourth-order term, needed to interpret the new experimental results, required evaluating 891 Feynman diagrams (*9*). This task involved numerical integrations on supercomputers over a period of more than 10 years, augmented by delicate analytical calculations that were required to deal with the infinities that underlie QED.

If the fine structure constant were known to a precision of 0.7 parts per billion, it could be inserted in the theoretical formula to provide a true test of QED. A discrepancy would be of major importance because it would be an indication of new physics. A number of different experiments have yielded values of α, but none with the precision required for this test. Consequently, the theoretical results are most usefully applied to extract a new value of α from the experiment. The new value is approximately 10 times as accurate as previous values. For the record, the value (expressed as an inverse value) found by Gabrielse and Kinoshita and their colleagues is $\alpha^{-1} = 137.035999710$, with an uncertainty of 0.7 parts per billion.

Although theories in physics all have boundaries to their areas of validity, nobody knows where that boundary is for QED. It is hoped that other measurements of α will continue to improve so that they can be combined with these new measurements to extend QED's area of validity or, better yet, find its boundary. Furthermore, there are a number of avenues for improving the measurements made by Gabrielse and his colleagues. The electron's magnetic moment is now known to better than a part per trillion, but the ultimate precision is not yet in sight.

References

1. B. Odom, D. Hanneke, B. D'Urso, G. Gabrielse, *Phys. Rev. Lett*. **97**, 030801 (2006).
2. G. Gabrielse, D. Hanneke, T. Kinoshita, M. Nio, B. Odom, *Phys. Rev. Lett.* **97**, 030802 (2006).
3. P. A. M. Dirac, *Proc. R. Soc. London A* **117**, 610 (1928).
4. P. A. M. Dirac, *Proc. R. Soc. London A* **118**, 351 (1928).
5. S. Schweber, *Q.E.D. and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga* (Princeton Univ. Press, Princeton, NJ, 1994).
6. P. Kusch, H. M. Foley, *Phys. Rev.* **74**, 250 (1948).
7. R. S. Van Dyck Jr., P. B. Schwinberg, H. G. Dehmelt, *Phys. Rev. Lett.* **59**, 26 (1987).
8. S. Laporta, E. Remiddi, *Phys. Lett. B* **379**, 283 (1996).
9. T. Kinoshita, M. Nio, *Phys. Rev. D* **73**, 013003 (2006).

10.1126/science.1131834

CELL SIGNALING

Protein Kinases Seek Close Encounters with Active Genes

John W. Edmunds and Louis C. Mahadevan

Signaling kinases may form integral components of transcription complexes, influencing gene expression in an unexpected way.

Upon exposure to changes in the environment or to developmental cues during differentiation, a cell reprograms transcription in its nucleus through a circuitry of signals that ultimately alters gene expression. Many of the steps of such signal-transducing cascades are executed by kinases, enzymes that transfer phosphate molecules onto target substrates. Often, kinases at the end of such cascades (terminal kinases) trigger the necessary response by directly phosphorylating transcription factors, coregulatory proteins, or the proteins that, with DNA, make up chromatin. Until recently, the prevailing view has been that terminal kinases operate enzymatically, without stable association with the chromatin that harbors target genes of a signaling pathway. But an alternative model whereby such kinases also play a structural role by binding to factors within transcription complexes at target genes has been slowly gathering support (*1*). On page 533 of this issue, Pokholok *et al*. (*2*) report a global analysis in yeast of the association of kinases with genes that they regulate, further supporting this model. Their findings suggest that such interactions can be observed not only with sequence-specific transcription factors positioned at regulatory (promoter) regions lying upstream of target genes, but also with the coding region of genes in some cases.

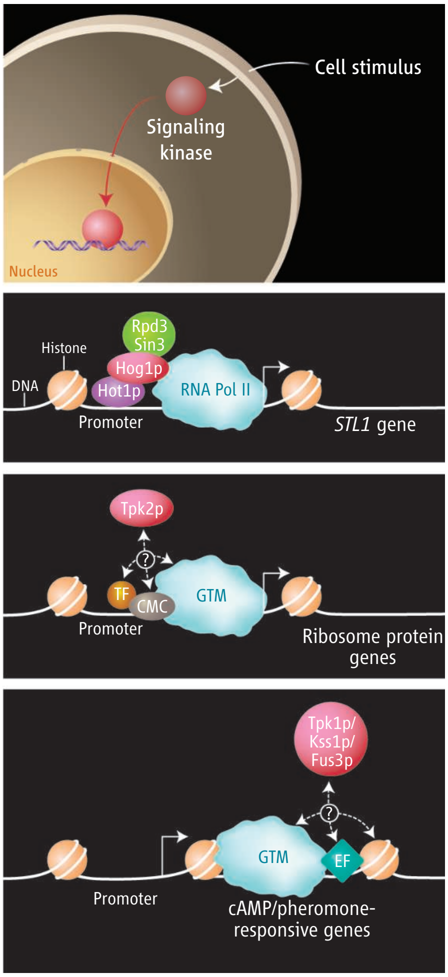
The yeast HOG mitogen-activated protein kinase (MAPK) pathway responds to changes in external osmolarity by activating the Hog1p MAPK, which then regulates expression of osmoresponsive genes (*3*, *4*). The necessity of its transcription factor substrate to retain Hog1p in the nucleus after cellular exposure to osmotic stress suggested that Hog1p might form stable interactions with its substrates, and experiments that identified potential binding partners for Hog1p indicated the same (*5*, *6*). A breakthrough came when chromatin immunoprecipitation (ChIP) experiments showed that in response to osmotic stress, Hog1p is

The authors are at the Nuclear Signalling Laboratory, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK. E-mail: louis.mahadevan@bioch.ox.ac.uk

recruited to particular target genes by transcription factors (*7*–*8*). Further work showed that Hog1p not only functions as a kinase at such genes, but also forms an integral component of transcription complexes involved in the recruitment of transcription factors, components of the general transcription machinery, RNA polymerase II (Pol II), and chromatin remodeling/modifying activities (*7*–*10*). This opened up the possibility that terminal kinases might have dual functions: a structural role, by mediating crucial protein-protein interactions within various transcription complexes, and an enzymatic role, by phosphorylating target proteins in such complexes to turn them on or off (*1*). Indeed, the finding that p38 MAPK—the mammalian homolog of Hog1p—associates with RNA Pol II (*9*) and also with the enhancer region of muscle-specific genes during myogenic differentiation (*11*) supports this model. Furthermore, MSK1/2, the kinase that p38 MAPK phosphorylates and activates in mammals, is a nuclear kinase that phosphorylates proteins associated with chromatin, including histone H3 and CREB (3′,5′-cyclic adenosine monophosphate response element–binding protein) (*12*–*13*). The MSK1/2-related kinase in *Drosophila melanogaster*, Jil-1, is reported to be chromatin associated (*14*). Thus, the physical and functional association of Hog1p/p38 MAPK with chromatin is quite well established. What about other gene-regulatory kinases?

Pokholok *et al.* extend this concept to other such kinases and a greater multitude of genes by combining the ChIP assay with DNA microarrays—so called "ChIP-on-chip" technology. The authors expand the subset of genes known to bind Hog1p in response to osmotic stress from 7 to 39, and they use a mutant yeast strain devoid of Hog1p to show that normal expression of most of these genes requires Hog1p. Binding is highest at the promoter region of these genes but is also detectable to a lesser extent at coding regions. Curiously, only 39 genes were found in this study (an array spanning 85% of the yeast genome), even though there are ~600 Hog1p-controlled osmoresponsive genes (*15*–*17*). Thus, perhaps only a subset of Hog1p-regulated genes requires Hog1p to stably bind to chromatin.

Pokholok *et al.* also show that Fus3p and Kss1p, kinases of the mating pheromone signaling pathway, physically associate with the coding regions of eight pheromone-responsive genes. Strikingly, the scaffold protein Ste5p, which interacts with Fus3p at the cell membrane, occupies the same gene coding regions, which suggests that adaptor proteins might be involved at specific genes in the indirect recruitment of additional factors by kinases. Finally, the authors show that the different catalytic subunits of protein kinase A (Tpk1p and Tpk2p) associate with particular genes. Tpk1p associates with the coding regions of most actively transcribed genes of yeast under normal conditions. Furthermore, the amount of Tpk1p binding to chromatin positively correlates with the transcription rate of the target genes. Loss of Tpk1p binding was observed when particular genes were repressed (increased Tpk1p binding was observed when these genes were activated). Tpk2p was observed largely at the promoter region of genes encoding ribosomal proteins, and this enrichment did not correlate with gene activity.



Kinase recruitment. (**First panel**) In response to cellular stimuli, some kinases are recruited to target genes. (**Second panel**) Hog1p is recruited by a transcription factor (Hot1p) to the promoter region of the *STL1* osmoresponsive target gene. Hog1p then recruits RNA Pol II and a histone deacetylase complex (Rpd3-Sin3) to control gene expression. (**Third and fourth panels**) The Tpk2p catalytic subunit of protein kinase A (PKA) is recruited to the promoter region of target genes, whereas the Tpk1p PKA catalytic subunit, Fus3p, and Kss1p are recruited to the coding regions. Although the mechanism and purpose of recruitment of such kinases are not known, they may involve factors that share similar intragenic locations. CMC, chromatin modifying complex; GTM, general transcription machinery; TF, transcription factor; EF, elongation factor.

CREDIT: P. HUEY/SCIENCE

This study raises several interesting issues. One quantitative aspect that deserves comment is the difference in the relative enrichment of chromatin-associated factors as determined through ChIP-based analysis. The enrichment varies from about 40× for the transcription factor Gcn4p to about 10× or less for the Hog1p and Tpk1p kinases (*2*). If all other experimental variables during ChIP experiments [such as antibody recovery differences (*18*)] are accounted for, this variation may indicate that the residence times of these proteins at these locations differ. For example, a stable interaction between a transcription factor and its target DNA is expected to give a higher recovery in ChIP-based analysis of the promoter region of a gene than the transient interaction of RNA Pol II at the coding region of the gene would recover coding sequences. Interpretation of quantitative differences in recovery by ChIP assays is fraught with complications but is unavoidable if we are to extract the full value of these data (*18*).

Differences in the types of genes and regions of genes with which these different kinases bind may reflect the mechanisms by which they are recruited and/or the functions that they carry out. For example, Hog1p localizes mostly to the promoter region of genes, where we would expect to find specific transcription factors, transcription initiation factors, and promoter-associated coregulatory proteins. This provides an obvious mechanism of protein-protein interaction for the specific recruitment of kinases. Previous findings have shown Hog1p to be recruited by promoter-bound transcription factors and that it functions in the recruitment of RNA Pol II (*7*–*9*). Similarly, Pokholok *et al.* show good correlation between the genic locations of Tpk2p, the Rap1p transcription factor, and the Esa1p subunit of the NuA4 chromatin-modifying complex (*2*). Thus, one could speculate that Rap1p recruits Tpk2p and/or Tpk2p aids in the recruitment of the NuA4 complex.

Less obvious with respect to mechanism is the finding of a correlation between the genic distribution of Tpk1p with RNA Pol II and specific histone H3 posttranslational modifications at the coding regions of some genes (*2*). There is no clear evidence that Tpk1p binds directly to posttranslationally modified histone tails at active genes. One speculation is that RNA Pol II and transcription are involved in the recruitment of Tpk1p to specific genes. This idea is supported by the positive correlation between transcription rate and Tpk1p gene association; if true, it raises the question of how Tpk1p is recruited specifically to particular genes and not to others that are being simultaneously transcribed by RNA Pol II. The presence of Hog1p in the coding regions of specific genes is easier to explain as Hog1p is also recruited to the promoters of these genes, and perhaps enters the coding regions by

Published by AAAS

"piggybacking" with RNA Pol II. Nonetheless, in this important study, Pokholok *et al*. widen the circumstances in which kinases may be found as a relatively stable constituent of chromatin at both promoter and coding regions of active genes. This may be a more widespread and general phenomenon than is currently appreciated.

References

1. J. W. Edmunds, L. C. Mahadevan, *J. Cell Sci.* **117**, 3715 (2004).
2. D. K. Pokholok, J. Zeitlinger, N. M. Hannett, D. B. Reynolds, R. A. Young, *Science* **313**, 533 (2006).
3. E. de Nadal, P. M. Alepuz, F. Posas, *EMBO Rep.* **3**, 735 (2002).
4. S. M. O'Rourke, I. Herskowitz, E. K. O'Shea, *Trends Genet.* **18**, 405 (2002).
5. M. Rep *et al.*, *Mol. Cell. Biol.* **19**, 5474 (1999).
6. V. Reiser, H. Ruis, G. Ammerer, *Mol. Biol. Cell* **10**, 1147 (1999).
7. M. Proft, K. Struhl, *Mol. Cell* **9**, 1307 (2002).
8. P. M. Alepuz, A. Jovanovic, V. Reiser, G. Ammerer, *Mol. Cell* **7**, 767 (2001).
9. P. M. Alepuz, E. de Nadal, M. Zapater, G. Ammerer, F. Posas, *EMBO J.* **22**, 2433 (2003).
10. E. de Nadal *et al.*, *Nature* **427**, 370 (2004).
11. C. Simone *et al.*, *Nat. Genet.* **36**, 738 (2004).
12. M. Deak, A. D. Clifton, L. M. Lucocq, D. R. Alessi, *EMBO J.* **17**, 4426 (1998).
13. A. Soloaga *et al.*, *EMBO J.* **22**, 2788 (2003).
14. Y. Wang, W. Zhang, Y. Jin, J. Johansen, K. M. Johansen, *Cell* **105**, 433 (2001).
15. S. M. O'Rourke, I. Herskowitz, *Mol. Biol. Cell* **15**, 532 (2003).
16. F. Posas *et al.*, *J. Biol. Chem.* **275**, 17249 (2000).
17. M. Rep, M. Krantz, J. M. Thevelein, S. Hohmann, *J. Biol. Chem.* **275**, 8290 (2000).
18. A. L. Clayton, C. A. Hazzalin, L. C. Mahadevan, *Mol. Cell*. in press.

10.1126/science.1131158

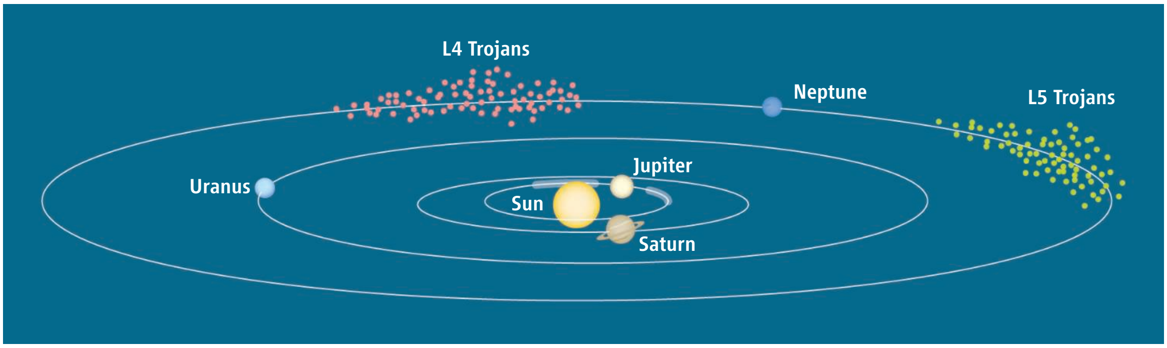
PLANETARY SCIENCE

Puzzling Neptune Trojans

An asteroid has been found in a highly inclined path co-orbiting with Neptune. Its discovery may help explain the evolution of the outer solar system.

Francesco Marzari

Trojan asteroids are small bodies that revolve about the Sun at the same distance as their host planet and share the planet's orbital path. They are locked at the two gravitationally stable locations, called triangular Lagrangian points, in distinct clouds that lead or trail the planet by about 60° (see the figure). Jupiter has the most of these Trojans, which are small rocky-icy bodies with diameters less than 300 km and are similar in composition to other minor bodies such as short-period comets, Kuiper Belt objects (KBOs), and Centaurs, small bodies that orbit between Jupiter and Neptune. About 2000 Jupiter Trojans are known today, but astronomers believe there may be as many of these asteroids in the kilometer-size range as there are main-belt asteroids (*1*). Four asteroids are also known to orbit in the Lagrangian points for Mars; these might possibly be rare remnants of planetesimals that formed in the terrestrial planet region. Moreover, Trojans are now known to gather near Neptune, and on page 511 of this issue, Sheppard and Trujillo report the discovery of the fourth such object (*2*), with important implications for theories of solar system formation.



Unusual asteroids. Trojan asteroids, small bodies that co-orbit with a planet in stable leading or trailing locations, are known to accompany Jupiter. They have also been discovered near Neptune, and Sheppard and Trujillo have now identified one with a highly inclined orbit.

CREDIT: P. HUEY/SCIENCE

Scientists theorize that Trojans are pristine bodies that originated very early in the history of the solar system and were captured in the final phase of planet formation. Different theories, not necessarily mutually exclusive, have been proposed to explain how planetesimals passing close to a planet fall into the force traps around the Lagrangian points. Among these are broadening of the tadpole-shaped regions of stable Trojan motion around the triangular Lagrangian points because of the growth of the planet's mass, direct collisional placement, drag-driven capture in the presence of the gaseous nebula, and chaotic trapping during giant planet migration (see below). There is as yet no general consensus on the source region of putative Trojans in the planetesimal disk. Some capture mechanisms demand that they formed near the planet's orbit, thus reflecting the physical and chemical composition of the planetary building blocks. The recent theory of chaotic capture, suggesting that planetesimals in temporary Trojan trajectories can be frozen into stable orbits as soon as planetary migration drives the host planet far away from a dynamically perturbed region (*3*), opens the possibility that Trojans might have formed in more distant regions of the planetesimal disk of the early solar system, sharing the same environment as KBOs.

In the course of the Deep Ecliptic Survey, a NASA-funded survey of the outer solar system, astronomers announced in 2001 the discovery of the first known member of a long-sought population of bodies: the Neptune Trojans. Sheppard and Trujillo report the discovery of the fourth object in this group, which is noteworthy in that it exhibits a high inclined orbit (about 25°). This finding strongly supports the idea that Neptune Trojans fill a thick disk with a population comparable to, or even larger than, that of Jupiter Trojans. At the same time, the discovery puts constraints on the mechanism by which they were captured.

What makes the Neptune Trojans so special for astronomers? According to recent theories, the outer solar system might have been a tumultuous environment. During the last stage of planetary formation, the giant planets may have migrated away from their formation sites by exchanging angular momentum with the residual planetesimal disk. Jupiter drifted inward, although only slightly, whereas Saturn, Uranus, and Neptune migrated outward by larger amounts. This past planetary migration explains many of the observable characteristics of KBOs, in particular of the resonant ones called Plutinos. However, the migration

The author is in the Department of Physics, University of Padova, Via Marzolo 8, Padova I-35131, Italy. E-mail: francesco.marzari@pd.infn.it

Published by AAAS

process may not have been so smooth as initially thought, and numerical simulations performed by Tsiganis *et al*. (*4*) show that the passage of Jupiter and Saturn through a 2:1 resonance may have ignited a period of strong chaotic evolution of Uranus and Neptune. In this scenario, the two planets had frequent close encounters and may even have exchanged orbits before their eccentricities finally settled down, allowing a more quiet migration to the present orbits.

The presence of a thick disk of Trojans around Neptune is clearly relevant to understanding the dynamical evolution of the planet. The co-orbital Trojan paths are unstable when Neptune has repeated close approaches with Uranus, and the capture of the present population appears possible either at the time of the last radial jump related to an encounter with Uranus or during the final period of slow migration. In this last case, collisional emplacement—in synergy with the reduction of the libration amplitude attributable to the outward migration and by the mass growth of the planet—is the only viable mechanism for trapping Trojans in this phase, but it does not appear to be so efficient as to capture a large population. Moreover, the only frequent planetesimal collisions are those that are close to the median plane of the disk, and this fact is at odds with the presence of high-inclination Trojans such as the one found by Sheppard and Trujillo. A thick disk of Neptune Trojans seems also to rule out the possibility that Trojans formed in situ from debris of collisions that occurred nearby (*5*).

The chaotic capture invoked to explain the orbital distribution of Jupiter Trojans might have worked out in the same way for Neptune. The planet at present is close to a 2:1 mean-motion resonance with Uranus; however, the resonance crossing has not been reproduced so far in numerical simulations of the migration of the outer planets. Alternatively, some sweeping secular resonance might have provided the right amount of instability for the "freeze-in" trapping to occur. In the near future, after additional Neptune Trojans are detected, an important test would be to look for a possible asymmetry between the trailing and leading clouds. Theoretical studies have shown that the L5 Lagrangian point (the trailing one) is more stable in the presence of outward radial migration and that this asymmetry strongly depends on the migration rate. This finding would have direct implications for the capture mechanism and for the possibility that the outward migration of Neptune was indeed smooth, without fast jumps caused by gravitational encounters with Uranus.

Sheppard and Trujillo also sort out another aspect of the known Neptune Trojans: their optical color distribution. It appears to be homogeneous and similar to that of Jupiter Trojans, irregular satellites, and possibly comets, but is less consistent with the color distribution of KBOs as a group. This finding raises questions about the compositional gradient along the planetesimal disk in the early solar system, the degree of radial mixing caused by planetary stirring, and the origin of the Jupiter and Neptune Trojans. Did Trojans form in a region of the planetesimal disk thermally and compositionally separated from that of the KBOs? How far did the initial solar nebula extend to allow important differences among small-body populations? Additional data are needed to solve the puzzles of the dynamical and physical properties of Neptune Trojans, and the finding by Sheppard and Trujillo is only the first step.

References

1. D. C. Jewitt, C. A. Trujillo, J. X. Luu, *Astron. J.* **120**, 1140 (2000).
2. S. S. Sheppard, C. A. Trujillo, *Science* **313**, 511 (2006); published online 15 June 2006 (10.1126/science.1127173).
3. A. Morbidelli, H. F. Levison, K. Tsiganis, R. Gomes, *Nature* **435**, 462 (2005).
4. K. Tsiganis, R. Gomes, A. Morbidelli, H. F. Levison, *Nature* **435**, 459 (2005).
5. E. I. Chiang, Y. Lithwick, *Astrophys. J.* **628**, L520 (2005).

Published online 15 June 2006;
10.1126/science.1129458
Include this information when citing this paper.

CLIMATE CHANGE

Can We Detect Trends in Extreme Tropical Cyclones?

Subjective measurements and variable procedures make existing tropical cyclone databases insufficiently reliable to detect trends in the frequency of extreme cyclones.

Christopher W. Landsea, Bruce A. Harper, Karl Hoarau, John A. Knaff

Recent studies have found a large, sudden increase in observed tropical cyclone intensities, linked to warming sea surface temperatures that may be associated with global warming (*1*–*3*). Yet modeling and theoretical studies suggest only small anthropogenic changes to tropical cyclone intensity several decades into the future [an increase on the order of ~5% near the end of the 21st century (*4*, *5*)]. Several comments and replies (*6*–*10*) have been published regarding the new results, but one key question remains: Are the global tropical cyclone databases sufficiently reliable to ascertain long-term trends in tropical cyclone intensity, particularly in the frequency of extreme tropical cyclones (categories 4 and 5 on the Saffir-Simpson Hurricane Scale)?

Tropical cyclone intensity is defined by the maximum sustained surface wind, which occurs in the eyewall of a tropical cyclone over an area of just a few dozen square kilometers. The main method globally for estimating tropical cyclone intensity derives from a satellite-based pattern recognition scheme known as the Dvorak Technique (*11*–*13*). The Atlantic basin has had routine aircraft reconnaissance since the 1940s, but even here, satellite images are heavily relied upon for intensity estimates, because aircraft can monitor only about half of the basin and are not available continuously. However, the Dvorak Technique does not directly measure maximum sustained surface wind. Even today, application of this technique is subjective, and it is common for different forecasters and agencies to estimate significantly different intensities on the basis of identical information.

The Dvorak Technique was invented in 1972 and was soon used by U.S. forecast offices, but the rest of the world did not use it routinely until the early 1980s (*11*, *13*). Until then, there was no systematic way to estimate the maximum sustained surface wind for most tropical cyclones. The Dvorak Technique was first developed for visible imagery (*11*), which precluded obtaining tropical cyclone intensity estimates at night and limited the sampling of maximum sustained surface wind. In 1984, a quantitative infrared method (*12*) was published, based on the observation that the temperature contrast between the warm eye of the cyclone and the cold cloud tops of the eyewall was a reasonable proxy for the maximum sustained surface wind.

In 1975, two geostationary satellites were available for global monitoring, both with 9-km resolution for infrared imagery. Today, eight

C. W. Landsea is at the NOAA National Hurricane Center, Miami, FL 33165, USA. E-mail: chris.landsea@ noaa.gov B. A. Harper is with Systems Engineering Australia Pty. Ltd., Bridgeman Downs, Queensland 4035, Australia. K. Hoarau is at the Cergy-Pontoise University, 95011 Cergy-Pontoise Cedex, France. J. A. Knaff is at the NOAA Cooperative Institute for Research in the Atmosphere, Fort Collins, CO 80523, USA.

Published by AAAS

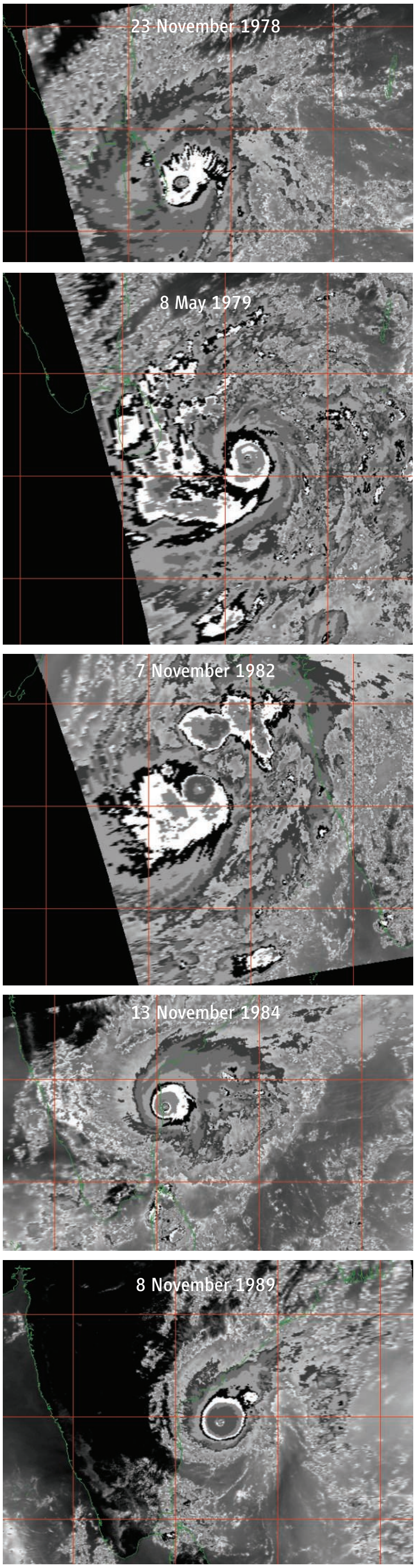
satellites are available with typically 4-km resolution in the infrared spectrum. The resulting higher resolution images and more direct overhead views of tropical cyclones result in greater and more accurate intensity estimates in recent years when using the infrared Dvorak Technique. For example (*13*), Atlantic Hurricane Hugo was estimated to have a maximum sustained surface wind of 59 m s^{-1} on 15 September 1989, based on use of the Dvorak Technique from an oblique observational angle. But in situ aircraft reconnaissance data obtained at the same time revealed that the hurricane was much stronger (72 m/s) than estimated by satellite. This type of underestimate was probably quite common in the 1970s and 1980s in all tropical cyclone basins because of application of the Dvorak Technique in an era of few satellites with low spatial resolution.

Operational changes at the various tropical cyclone warning centers probably also contributed to discontinuities in tropical cyclone intensity estimates and to more frequent identification of extreme tropical cyclones (along with a shift to stronger maximum sustained surface wind in general) by 1990. These operational changes include (*13*–*17*) the advent of advanced analysis and display systems for visualizing satellite images, changes in the pressure-wind relationships used for wind estimation from observed pressures, relocation of some tropical cyclone warning centers, termination of aircraft reconnaissance in the Northwest Pacific in August 1987, and the establishment of specialized tropical cyclone warning centers.

Therefore, tropical cyclone databases in regions primarily dependent on satellite imagery for monitoring are inhomogeneous and likely to have artificial upward trends in intensity. Data from the only two basins that have had regular aircraft reconnaissance—the Atlantic and Northwest Pacific—show that no significant trends exist in tropical cyclone activity when records back to at least 1960 are examined (*7*, *9*). However, differing results are obtained if large bias corrections are used on the best track databases (*1*), although such strong adjustments to the tropical cyclone intensities may not be warranted (*7*). In both basins, monitoring and operational changes complicate the identification of true climate trends. Tropical cyclone "best track" data sets are finalized annually by operational meteorologists, not by climate researchers, and none of the data sets have been quality controlled to account for changes in physical understanding, new or modified methods for analyzing intensity, and aircraft/satellite data changes (*18*–*21*).

To illustrate our point, the figure presents satellite images of five tropical cyclones listed in the North Indian basin database for the period 1977 to 1989 as category 3 or weaker. Today, these storms would likely be considered extreme tropical cyclones based on retrospective application of the infrared Dvorak Technique. Another major tropical cyclone, the 1970 Bangladesh cyclone—the world's worst tropical-cyclone disaster, with 300,000 to 500,000 people killed—does not even have an official intensity estimate, despite indications that it was extremely intense (*22*). Inclusion of these storms as extreme tropical cyclones would boost the frequency of such events in the 1970s and 1980s to numbers indistinguishable from the past 15 years, suggesting no systematic increase in extreme tropical cyclones for the North Indian basin.

Previously unrecognized category 4 and 5 tropical cyclones in the North Indian Ocean, 1978–1989



Underestimated storm intensity. The North Indian basin tropical cyclones shown here are listed in the best track data set as category 3 or weaker, but were probably category 4 or 5. Similar underestimates may have been common in all ocean basins in the 1970s and 1980s. Trend analyses for tropical cyclones intensities are therefore highly problematic.

These examples are not likely to be isolated exceptions. Ongoing Dvorak reanalyses of satellite images in the Eastern Hemisphere basins by the third author suggest that there are at least 70 additional, previously unrecognized category 4 and 5 cyclones during the period 1978–1990. The pre-1990 tropical cyclone data for all basins are replete with large uncertainties, gaps, and biases. Trend analyses for extreme tropical cyclones are unreliable because of operational changes that have artificially resulted in more intense tropical cyclones being recorded, casting severe doubts on any such trend linkages to global warming.

There may indeed be real trends in tropical cyclone intensity. Theoretical considerations based on sea surface temperature increases suggest an increase of ~4% in maximum sustained surface wind per degree Celsius (*4*, *5*). But such trends are very likely to be much smaller (or even negligible) than those found in the recent studies (*1*–*3*). Indeed, Klotzbach has shown (*23*) that extreme tropical cyclones and overall tropical cyclone activity have globally been flat from 1986 until 2005, despite a sea surface temperature warming of 0.25°C. The large, step-like increases in the 1970s and 1980s reported in (*1*–*3*) occurred while operational improvements were ongoing. An actual increase in global extreme tropical cyclones due to warming sea surface temperatures should have continued during the past two decades.

Efforts under way by climate researchers—including reanalyses of existing tropical cyclone databases (*20*, *21*)—may mitigate the problems in applying the present observational tropical cyclone databases to trend analyses to answer the important question of how humankind may (or may not) be changing the frequency of extreme tropical cyclones.

References and Notes

1. K. Emanuel, *Nature* **436**, 686 (2005).
2. P. J. Webster, G. J. Holland, J. A. Curry, H.-R. Chang, *Science* **309**, 1844 (2005).
3. C. D. Hoyos, P. A. Agudelo, P. J. Webster, J. A. Curry,

Science **312**, 94 (2006); published online 15 March 2006 (10.1126/science.1123560).
4. T. R. Knutson, R. E. Tuleya, *J. Clim.* **17**, 3477 (2004).
5. K. Emanuel, in *Hurricanes and Typhoons: Past, Present and Future*, R. J. Murnane, K.-B. Liu, Eds. (Columbia Univ. Press, New York, 2004), pp. 395–407.
6. R. A. Pielke Jr., *Nature* **438**, E11 (2005).
7. C. W. Landsea, *Nature* **438**, E11 (2005).
8. K. Emanuel, *Nature* **438**, E13 (2005).
9. J. C. L. Chan, *Science* **311**, 1713b (2006).
10. P. J. Webster, J. A. Curry, J. Liu, G. J. Holland, *Science* **311**, 1713c (2006).
11. V. F. Dvorak, *Mon. Weather Rev.* **103**, 420 (1975).
12. V. F. Dvorak, *NOAA Tech. Rep. NESDIS 11* (1984).
13. C. Velden *et al.*, *Bull. Am. Meteorol. Soc.*, in press.
14. J. A. Knaff, R. M. Zehr, *Weather Forecast.*, in press.
15. C. Neumann, *in Storms Volume 1*, R. Pielke Jr., R. Pielke Sr., Eds. (Routledge, New York, 2000), pp. 164–195.
16. R. J. Murnane, in *Hurricanes and Typhoons: Past, Present and Future*, R. J. Murnane, K.-B. Liu, Eds. (Columbia Univ. Press, New York, 2004), pp. 249–266.
17. J.-H. Chu, C. R. Sampson, A. S. Levine, E. Fukada, *The Joint Typhoon Warning Center Tropical Cyclone Best-Tracks, 1945–2000*, Naval Research Laboratory Reference Number NRL/MR/7540-02-16 (2002).
18. C. W. Landsea, *Mon. Weather Rev.* **121**, 1703 (1993).
19. J. L. Franklin, M. L. Black, K. Valde, *Weather Forecast.* **18**, 32 (2003).
20. C. W. Landsea *et al.*, *Bull. Am. Meteorol. Soc.* **85**, 1699 (2004).
21. C. W. Landsea *et al.*, in *Hurricanes and Typhoons: Past, Present and Future*, R. J. Murnane, K.-B. Liu, Eds. (Columbia Univ. Press, New York, 2004), pp. 177–221.
22. K. Emanuel, *Divine Wind—The History and Science of Hurricanes* (Oxford Univ. Press, Oxford, 2005).
23. P. J. Klotzbach, *Geophys. Res. Lett.* **33**, 10.1029/2006GL025881 (2006).

24. This work was sponsored by a grant from the NOAA Climate and Global Change Program on the Atlantic Hurricane Database Re-analysis Project. Helpful comments and suggestions were provided by L. Avila, J. Beven, E. Blake, J. Callaghan, J. Kossin, T. Knutson, M. Mayfield, A. Mestas-Nunez, R. Pasch, and M. Turk.

10.1126/science.1128448

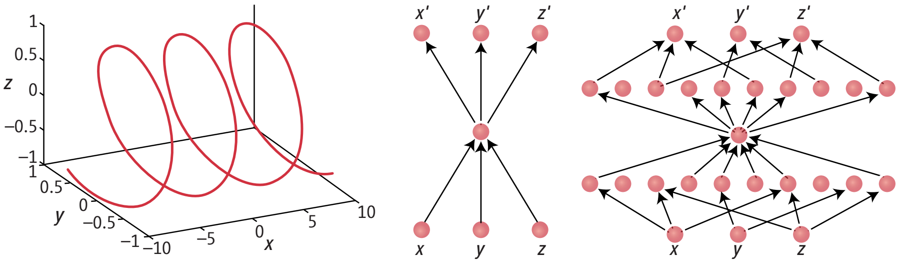
COMPUTER SCIENCE

New Life for Neural Networks

With the help of neural networks, data sets with many dimensions can be analyzed to find lower dimensional structures within them.

Garrison W. Cottrell

As many researchers have found, the data they have to deal with are often high-dimensional—that is, expressed by many variables—but may contain a great deal of latent structure. Discovering that structure, however, is nontrivial. To illustrate the point, consider a case in the relatively low dimension of three. Suppose you are handed a large number of three-dimensional points in random order (where each point is denoted by its coordinates along the x, y, and z axes): $\{(-7.4000, -0.8987, 0.4385), (3.6000, -0.4425, -0.8968), (-5.0000, 0.9589, 0.2837), \ldots\}$. Is there a more compact, lower dimensional description of these data? In this case, the answer is yes, which one would quickly discover by plotting the points, as shown in the left panel of the figure. Thus, although the data exist in three dimensions, they really lie along a one-dimensional curve that is embedded in three-dimensional space. This curve can be represented by three functions of x, as $(x, y, z) = [x, \sin(x), \cos(x)]$. This immediately reveals the inherently one-dimensional nature of these data. An important feature of this description is that the natural distance between two points is not the Euclidean, straight line distance; rather, it is the distance along this curve. As Hinton and Salakhutdinov report on page 504 of this issue ([1]), the discovery of such low-dimensional encodings of very high-dimensional data (and the inverse transformation back to high dimensions) can now be efficiently carried out with standard neural network techniques. The trick is to use networks initialized to be near a solution, using unsupervised methods that were recently developed by Hinton's group.



Searching for structure. (**Left**) Three-dimensional data that are inherently one-dimensional. (**Middle**) A simple "autoencoder" network that is designed to compress three dimensions to one, through the narrow hidden layer of one unit. The inputs are labeled x, y, z, with outputs x', y', and z'. (**Right**) A more complex autoencoder network that can represent highly nonlinear mappings from three dimensions to one, and from one dimension back out to three dimensions.

This low-dimensional structure is not uncommon; in many domains, what initially appears to be high-dimensional data actually lies upon a much lower dimensional manifold (or surface). The issue to be addressed is how to find such lower dimensional descriptions when the form of the data is unknown in advance, and is of much higher dimension than three. For example, digitized images of faces taken with a 3-megapixel camera exist in a very high dimensional space. If each pixel is represented by a gray-scale value between 0 and 255 (leaving out color), the faces are points in a 3-million-dimensional hypercube that also contains all gray-scale pictures of that resolution. Not every point in that hypercube is a face, however, and indeed, most of the points are not faces. We would like to discover a lower dimensional manifold that corresponds to "face space," the space that contains all face images and only face images. The dimensions of face space will correspond to the important ways that faces differ from one another, and not to the ways that other images differ.

This problem is an example of unsupervised learning, where the goal is to find underlying regularities in the data, rather than the standard supervised learning task where the learner must classify data into categories supplied by a teacher. There are many approaches to this problem, some of which have been reported in this journal ([2], [3]). Most previous systems learn the local structure among the points—that is, they can essentially give a neighborhood structure around a point, such that one can measure distances between points within the manifold. A major limitation of these approaches, however, is that one cannot take a new point and decide where it goes on the underlying manifold ([4]). That is, these approaches only learn the underlying low-dimensional structure of a given set of data, but they do not provide a mapping from new data points in the high-dimensional space into the structure that they have found (an encoder), or, for that matter, a mapping back out again into the original space (a decoder). This is an important feature because without it, the method can only be applied to the original data set, and cannot be used on novel data. Hinton and Salakhutdinov address the issue of finding an invertible mapping by making a known but previously impractical

The author is in the Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA 92093–0404, USA. E-mail: gary@cs.ucsd.edu

Published by AAAS

method work effectively. They do this by making good use of recently developed machine learning algorithms for a special class of neural networks (*5*, *6*).

Hinton and Salakhutdinov's approach uses so-called autoencoder networks—neural networks that learn a compact description of data, as shown in the middle panel of the figure. This is a neural network that attempts to learn to map the three-dimensional data from the spiral down to one dimension, and then back out to three dimensions. The network is trained to reproduce its input on its output—an identity mapping—by the standard backpropagation of error method (*7*, *8*). Although backpropagation is a supervised learning method, by using the input as the teacher, this method becomes unsupervised (or self-supervised). Unfortunately, this network will fail miserably at this task, in much the same way that standard methods such as principal components analysis will fail. This is because even though there is a weighted sum of the inputs (a linear mapping) to a representation of x—the location along the spiral—there is no (semi-)linear function (*9*) of x that can decode this back to $\sin(x)$ or $\cos(x)$. That is, the network is incapable of even representing the transformation, much less learning it. The best such a network can do is to learn the average of the points, a line down the middle of the spiral. However, if another nonlinear layer is added between the output and the central hidden layer (see the figure, right panel), then the network is powerful enough, and can learn to encode the points as one dimension (easy) but also can learn to decode that one-dimensional representation back out to the three dimensions of the spiral (hard). Finding a set of connection strengths (weights) that will carry out this learning problem by means of backpropagation has proven to be unreliable in practice (*10*). If one could initialize the weights so that they are near a solution, it is easy to fine-tune them with standard methods, as Hinton and Salakhutdinov show.

The authors use recent advances in training a specific kind of network, called a restricted Boltzmann machine or Harmony network (*5*, *6*), to learn a good initial mapping recursively. First, their system learns an invertible mapping from the data to a layer of binary features. This initial mapping may actually increase the dimensionality of the data, which is necessary for problems like the spiral. Then, it learns a mapping from those features to another layer of features. This is repeated as many times as desired to initialize an extremely deep autoencoder. The resulting deep network is then used as the initialization of a standard neural network, which then tunes the weights to perform much better.

This makes it practical to use much deeper networks than were previously possible, thus allowing more complex nonlinear codes to be learned. Although there is an engineering flavor to much of the paper, this is the first practical method that results in a completely invertible mapping, so that new data may be projected into this very low dimensional space. The hope is that these lower dimensional representations will be useful for important tasks such as pattern recognition, transformation, or visualization. Hinton and Salakhutdinov have already demonstrated some excellent results in widely varying domains. This is exciting work with many potential applications in domains of current interest such as biology, neuroscience, and the study of the Web.

Recent advances in machine learning have caused some to consider neural networks obsolete, even dead. This work suggests that such announcements are premature.

References and Notes

1. G. E. Hinton, R. R. Salakhutdinov, *Science* **313**, 504 (2006).
2. S. T. Roweis, L. K. Saul, *Science* **290**, 2323 (2000).
3. J. A. Tenenbaum, V. J. de Silva, J. C. Langford, *Science* **290**, 2319 (2000).
4. One can learn a mapping to the manifold (and back), but this is done independently of the original structure-finding method, which does not provide this mapping.
5 G. E. Hinton, *Neural Comput.* **14**, 1771 (2002).
6. P. Smolensky, in *Parallel Distributed Processing*, vol. 1, *Foundations*, D. E. Rumelhart, J. L. McClelland, PDP Research Group, Eds. (MIT Press, Cambridge, MA, 1986), pp. 194–281.
7. D. E. Rumelhart, G. E. Hinton, R. J. Williams, *Nature* **323**, 533 (1986).
8. G. W. Cottrell, P. W. Munro, D. Zipser, in *Models of Cognition: A Review of Cognitive Science*, N. E. Sharkey, Ed. (Ablex, Norwood, NJ, 1989), vol. 1, pp. 208–240.
9. A so-called semilinear function is one that takes as input a weighted sum of other variables, and applies a monotonic transformation to it. The standard sigmoid function used in neural networks is an example.
10. D. DeMers, G. W. Cottrell, in *Advances in Neural Information Processing Systems*, S. J. Hanson, J. D. Cowan, C. L. Giles, Eds. (Morgan Kaufmann, San Mateo, CA, 1993), vol. 5, pp. 580–587.

10.1126/science.1129813

ATMOSPHERE

What Drives the Ice Age Cycle?

Didier Paillard

Between 3 and 1 million years ago, ice ages followed a 41,000-year cycle. Two studies provide new explanations for this periodicity.

The exposure of Earth's surface to the Sun's rays (or insolation) varies on time scales of thousands of years as a result of regular changes in Earth's orbit around the Sun (eccentricity), in the tilt of Earth's axis (obliquity), and in the direction of Earth's axis of rotation (precession). According to the Milankovitch theory, these insolation changes drive the glacial cycles that have dominated Earth's climate for the past 3 million years.

For example, between 3 and 1 million years before present (late Pliocene to early Pleistocene, hereafter LP-EP), the glacial oscillations followed a 41,000-year cycle. These oscillations correspond to insolation changes driven by obliquity changes. But during this time, precession-driven changes in insolation on a 23,000-year cycle were much stronger than the obliquity-driven changes. Why is the glacial record for the LP-EP dominated by obliquity, rather than by the stronger precessional forcing? How should the Milankovitch theory be adapted to account for this "41,000-year paradox"?

Two different solutions are presented in this issue. The first involves a rethinking of how the insolation forcing should be defined (*1*), whereas the second suggests that the Antarctic ice sheet may play an important role (*2*).The two papers question some basic principles that are often accepted without debate.

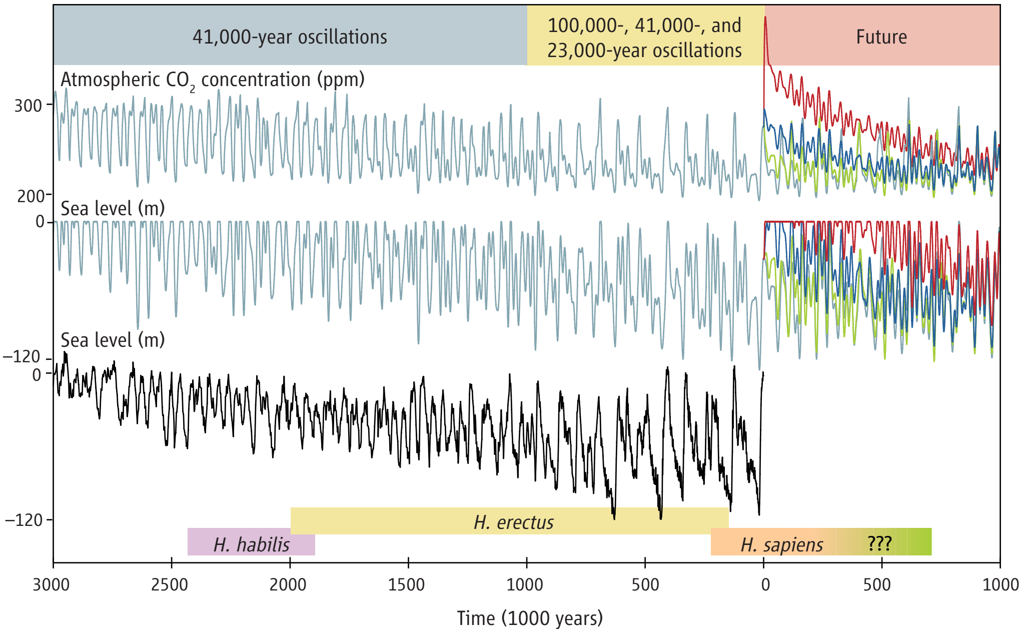
On page 508, Huybers (*1*) argues that the summer insolation traditionally used in ice age models may not be the best parameter. Because ice mass balance depends on whether the temperature is above or below the freezing point, a physically more relevant parameter should be the insolation integrated over a given threshold that allows for ice melting. This new parameter more closely follows a 41,000-year periodicity, thus providing a possible explanation for the LP-EP record.

On page 492, Raymo *et al*. (*2*) question another pillar of ice age research by suggesting that the East Antarctic ice sheet could have contributed substantially to sea-level changes during the LP-EP. The East Antarctic ice sheet is land-based and should therefore be sensitive mostly to insolation forcing, whereas the West Antarctic ice sheet is marine-based and thus influenced largely by sea-level changes. Because the obliquity forcing is symmetrical with respect to the hemispheres, whereas the preces-

The author is at the Laboratoire des Sciences du Climat et de l'Environnement, Institut Pierre Simon Laplace, CEA-CNRS-UVSQ, 91191 Gif-sur-Yvette, France. E-mail: didier.paillard@cea.fr

Published by AAAS

sional forcing is antisymmetrical, the contributions of the northern and southern ice sheets to the global ice volume record will add up for the 41,000-year cycle, but cancel each other out for the 23,000-year cycle, thus explaining the 41,000-year paradox.



Past and future climate. Simulated cycles of atmospheric CO_2 concentrations (**top**) and sea level (**middle**) from 3 million years before present to 1 million years in the future (*5*). The model accounts for the interaction between ice volume and atmospheric CO_2 concentrations. The amplitude of future climatic cycles may share similarities with those in the late Pliocene (about 3 million years ago), depending on the total amount of CO_2 released into the atmosphere through human activities (*8*). Gray: without anthropogenic CO_2 emissions; green: 450 gigatons of carbon (GtC), assuming that emissions stop today; blue: 1500 GtC, an optimistic emissions scenario; red: 5000 GtC, a pessimistic emissions scenario, assuming that the entire estimated reservoir of fossil fuels on Earth is burnt. (**Bottom**) Isotopic record of past ice volume, showing 41,000-year cycles between 3 and 1 million years ago and larger 100,000-year cycles since 1 million years ago (*9*).

Both hypotheses could be part of the solution. Huybers's idea is based on a sound and simple physical premise and is certainly valid to some extent. The hypothesis of Raymo *et al.* provides a scenario for an increasing contribution of the 23,000-year cycles under a colder climate, through a transition from a land-based to a marine-based East Antarctic ice sheet around 1 million years ago. Indeed, though not dominant, the precessional cycles are present in the climate record of the past 1 million years (the late Pleistocene). Still, neither hypothesis can account for the beginning of Northern Hemisphere glaciations around 3 million years ago. Furthermore, during the past 1 million years, glacial-interglacial oscillations have largely been dominated by a 100,000-year periodicity, yet there is no notable associated 100,000-year insolation forcing. There is currently no consensus on what drives these late Pleistocene 100,000-year cycles.

The theories of Huybers and Raymo *et al.* can be traced back to the 19th century. In 1842, Adhémar proposed that the ice ages were driven by precessional changes (obliquity and eccentricity changes were unknown at this time). Because precessional changes are antisymmetrical with respect to the hemispheres, he argued that Antarctica is glaciated today, whereas some time ago, the northern hemisphere was covered by ice, thus explaining the geologic field data (*3*). This alternation between the hemispheres is somewhat like in (*2*). His theory was dismissed at the time by Lyell and by Alexander von Humboldt (*3*), because the amount of energy received on Earth does not depend on precession: more intense (colder) winters were also shorter, with the energy budget at the top of the atmosphere being unchanged because precession modulates not only the intensity but also the duration of seasons. Precession should thus not affect climate, somewhat like in (*1*).

Since the 19th century, two families of ice age theories have been put forward: insolation-based theories proposed by Adhémar, Croll, and Milankovitch, and atmospheric CO_2 ones proposed by Tyndall, Arrhenius, and Chamberlin (*3*). The latter theories suggested that glaciations were associated with lower CO_2 levels. This is now confirmed by the large oscillations in atmospheric CO_2 measured in Antarctic ice cores over the past 650,000 years (*4*). It is certainly difficult to explain the ice ages of the past 1 million years purely on the basis of insolation changes. In the late Pleistocene, both insolation changes and atmospheric CO_2 concentrations must have played a critical role in the dynamics of glaciations, although a final synthesis still eludes us.

The big challenge is to build an ice age theory that can account not only for ice sheet and atmospheric CO_2 changes, but also for the start of glaciations about 3 million years ago and for the transition from 41,000-year cycles to much larger 100,000-year oscillations around 1 million years ago. The atmospheric CO_2 concentration was probably very important over the past 1 million years, but was this also the case during the LP-EP? Alternatively, if one can build a purely insolation-based theory between 3 and 1 million years ago, as suggested by Huybers and Raymo *et al.*, why is this not the case anymore in the past 1 million years?

A tentative scenario, based on a bistable ocean system (*5*), is shown in the figure, where the 41,000-year paradox and the 100,000-year problem have a common answer in an oceanic switch that can store or release carbon depending on ice-sheet size and insolation forcing, using empirical relationships. This conceptual model can be extrapolated to a future with and without anthropogenic CO_2 emissions. The results are comparable to those of more sophisticated models (*6*), providing a framework for understanding the likely climatic future of our planet in the context of the climate of the past 3 million years.

The mid-Pliocene, about 3.3 to 3.0 million years ago, has been cited as a possible analog for our future warmer Earth (*7*). This and the subsequent LP-EP time period are interesting not only in terms of their climate, but also because during this period, *Homo habilis* first appeared on the scene. Furthermore, they are currently our best guide to what climate and ice sheets may look like for *Homo sapiens* to come. The reports by Huybers and by Raymo *et al.* bring us a step closer to understanding the dynamics of these past climates.

References and Notes

1. P. Huybers, *Science* **313**, 508 (2006).
2. M. E. Raymo, L. E. Lisiecki, K. H. Nisancioglu, *Science* **313**, 492 (2006).
3. E. Bard, *C. R. Geosci.* **336**, 603 (2004).
4. U. Siegenthaler *et al.*, *Science* **310**, 1313 (2005).
5. D. Paillard, F. Parrenin, *Earth Planet. Sci. Lett.* **227**, 263 (2004).
6. D. Archer, A. Ganopolski, *Geochem. Geophys. Geosyst.* **6**, Q05003 (2005).
7. H. Dowsett *et al.*, *Global Planet. Change* **9**, 169 (1994).
8. To build this figure, the model in (*5*) was extrapolated using a decay *e*-folding time of 400,000 years for the removal by silicate weathering of a remaining 8% long-lived part of total anthropogenic carbon, following (*10*).
9. L. E. Lisiecki, M. E. Raymo, *Paleoceanography* **20**, PA1003 (2005).
10. D. Archer, *J. Geophys. Res.* **110**, C09S05 (2005).

10.1126/science.1131297

Published by AAAS

AAAS NEWS & NOTES EDITED BY EDWARD W. LEMPINEN

SCIENCE AND LAW

Neuroscience in the Courts— A Revolution in Justice?

New imaging tools that show the brain in action raise the prospect that the courts might someday be able to reliably assess whether a witness has lied during pre-trial statements or whether a candidate for probation has a propensity to violence. But if human actions ever could be explained by a close analysis of the firing of neurons, would a criminal defendant then be able to claim that he is not really guilty but simply the victim of a "broken brain"?

That is the sort of question judges and lawyers may have to grapple with in the courtroom in the future—and at a seminar organized by AAAS, 16 state and federal judges got an intriguing preview of the emerging issues. The seminar, held 29 to 30 June at the Dana Center in Washington, D.C., was co-sponsored by the Federal Judicial Center and the National Center for State Courts, with funding from the Charles A. Dana Foundation.

Barbara Jacobs Rothstein and David Heeger.

Experts told the judges about brain-scanning technologies such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET). They heard about the formation of memory and whether it may be possible to distinguish true memories from false ones. They also heard about the possible neurological bases for violent and antisocial behavior.

The judges broke into teams to consider several hypothetical case studies, including whether a brain scanner that proved capable of identifying a propensity to violence should be used in jail assignments for convicted felons or to help decide whether a job applicant is suitable for employment. In general, the judicial reaction was cautious, with much talk about how to define "propensity" and whether such judgments can ever be made in isolation.

There was lively discussion about fMRI, a technology that can produce real-time images of people's brains as they answer questions, listen to sounds, view images, and respond to other stimuli. Some studies have shown that several regions of the brain, including the anterior cingulate cortex, appear to be active when a person is lying. Two private companies already are marketing fMRI "lie detection" services to police departments and U.S. government agencies, including the Department of Defense, the Department of Justice, the National Security Agency, and the CIA.

But David Heeger, a professor of psychology and neural science at New York University, cautioned the judges that fMRI is not a suitable lie detector now and may never fill the bill, even though it has the potential to outperform the traditional polygraph. In key studies, research subjects were instructed to lie and tested in settings where they knew there would be no serious consequences for lying. Moreover, the anterior cingulate cortex and other brain areas implicated in lying appear to play roles in a wide range of cognitive functions. So it is difficult to draw a specific link between activity in these brain regions and lying, critics say.

Such issues are of more than academic interest to the judges. Under the U.S. Supreme Court's *Daubert* ruling in 1993 and two subsequent rulings, trial judges have a gatekeeping responsibility in determining the validity of scientific evidence and all expert testimony.

"We judges are often at a point where we have to make very important decisions at the cutting edge of the juxtaposition of law and science," said Barbara Jacobs Rothstein, director of the Federal Judicial Center and a federal judge for the Western District of Washington state.

As science gains a better understanding of the physical basis in the brain for certain behaviors, some specialists argue that concepts such as free will, competency, and legal responsibility may be open to challenge. Against that backdrop, they say, it is important that judges be educated and informed about the scientific status of such neuroscience methods as imaging studies.

"I think law generally is behind the curve of science," said Stephen Spindler, a state judge in Indiana. "We don't get to deal with these things until someone springs them upon us. Law is reactive, not proactive, and we're getting a preview of what we can expect, maybe not tomorrow or next year, but coming down the pike."

The judicial seminar continues the effort by AAAS to bring together specialists from diverse fields to talk about the implications of neuroscience. Mark S. Frankel, the head of AAAS's Scientific Freedom, Responsibility and Law Program, said another neuroscience seminar for judges will be held 7 to 8 December at Stanford University in California.

— *Earl Lane*

SCIENCE COMMUNICATION

Science, AAAS Assess "State of the Planet"

The ability to address the critical environmental issues of our time—such as climate change, the health of Earth's oceans, and sustainability—is often checked by uncertainty and misunderstanding among policy-makers and the public. Now *Science* and AAAS have published a new volume that is designed to provide a state-of-the-art assessment of the complex, interrelated challenges that will shape our environmental future.

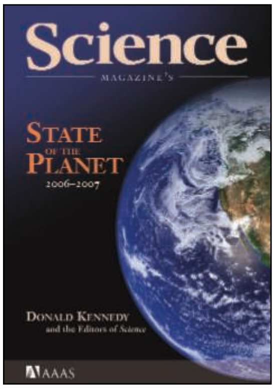
"*Science* Magazine's State of the Planet 2006–2007" [Island Press, June 2006, 201 pp.; $16.95 soft/$32 hard; ISBN: 1597260630] provides a clear, accessible view of scientific consensus on the environmental threats confronting Earth. The new volume includes three dozen essays and news stories, written by some of the world's most respected researchers, policy experts, and science journalists.

In the book's introduction, *Science* Editor-in-Chief Donald Kennedy notes that resources essential to life on Earth are closely connected to the health of the environment. The quality of fresh water depends on the condition of watershed forests. Agriculture depends on the vitality of surrounding ecosystems that are home to bees and birds. Climate change affects the distribution of plants and animals in the wild.

"To the editors of *Science*, these relationships—and the changes in them as humans continue to alter the world—comprise the most important and challenging issues societies

Published by AAAS

face,” Kennedy writes. “Without scientific understanding, those who will make policies in the future will be forced to do so without the most essential tool they could have.”



“*Science* Magazine’s State of the Planet 2006–2007”

The new book is a compilation of articles previously published in *Science* and recently updated, plus three new summary essays by Kennedy. The articles were chosen and assembled by editors at the journal.

At the heart of the book is a landmark 1968 essay in *Science,* “The Tragedy of the Commons,” by the late Garrett Hardin, formerly a professor of human ecology at the University of California at Santa Barbara. (“The Commons” is a term that describes the environment shared by all of life, and on which all of life depends.) Other essays in the new book originally were published in *Science* in November and December 2003 as part of a series called “The State of the Planet.”

The new book features an international roster of top environmental scholars. One of the essays, “The Struggle to Govern the Commons,” won the 2005 Sustainability Science Award from the Ecological Society of America. It was written by Thomas Dietz, director of the Environmental Science and Policy Program at Michigan State University; Elinor Ostrom, co-director of the Center for the Study of Institutions, Population and Environmental Change at Indiana University; and Paul C. Stern at the Division of Social and Behavioral Sciences and Education at the U.S. National Academies in Washington, D.C.

Among the other contributors:

- Martin Jenkins from the World Conservation Monitoring Centre of the United Nations Environment Programme in Cambridge, U.K., writing on the prospects for biodiversity. Jenkins is co-author of the “World Atlas of Biodiversity”;
- Hajime Akimoto, director of the Atmospheric Composition Research Program at the Frontier Research Center for Global Change in Yokohama, Japan, writing on global air quality;
- Robert T. Watson, chief scientist and director for Environmentally and Socially Sustainable Development at the World Bank, writing on climate change and the Kyoto Protocol; and
- Joel E. Cohen, an award-winning researcher, prolific author and head of the Laboratory of Populations at Rockefeller University and Columbia University in New York, writing on population.

To order the book, go to www.islandpress.org and search for “State of the Planet.”

SCIENCE POLICY

AAAS Testifies on Stem Cell Research

AAAS CEO Alan I. Leshner recommended to a U.S. Senate panel that federally funded science should explore the broadest possible range of stem cell research, including techniques that require the use of early-stage human embryos.

Leshner, the executive publisher of *Science,* was among those who testified on a bill co-sponsored by U.S. Senators Rick Santorum and Arlen Specter, both Pennsylvania Republicans, to promote stem cell research that does not require the use of human embryos.

Such research is important, Leshner told a subcommittee of the Appropriations Committee on 27 June. But, he added, the most promising avenues to date appear to be derivation of stem cells from early-stage embryos at in vitro fertilization (IVF) clinics or created by somatic cell nuclear transfer. “The alternatives that are now being developed are, in fact, intriguing,” Leshner said, “but we really don’t know what their ultimate utility will be, and each has potential problems or complications.”

Specter said he backs research on alternative stem cell methods, while continuing to push for a vote on legislation he has co-sponsored with Senator Tom Harkin (D–IA) that would authorize federally funded research on new stem cell lines derived from the microscopic embryos left over in the IVF process. President George W. Bush issued a directive in 2001 that federal dollars could be used for research only on embryonic stem cell lines already in existence at that time.

— Earl Lane

ELECTIONS

AAAS Annual Election: Preliminary Announcement

The 2006 AAAS election of general and section officers will be held in September. All members will receive a ballot for election of the president-elect, members of the Board of Directors, and members of the Committee on Nominations. Members registered in one to three sections will receive ballots for election of the chair-elect, member-at-large of the Section Committee, and members of the Electorate Nominating Committee for each section.

Members enrolled in the following sections will also elect Council delegates: Anthropology; Astronomy; Biological Sciences; Chemistry; Geology and Geography; Mathematics; Neuroscience; and Physics.

Candidates for all offices are listed below. Additional names may be placed in nomination for any office by petition submitted to the Chief Executive Officer no later than 25 August. Petitions nominating candidates for president-elect, members of the Board, or members of the Committee on Nominations must bear the signatures of at least 100 members of the Association. Petitions nominating candidates for any section office must bear the signatures of at least 50 members of the section. A petition to place an additional name in nomination for any office must be accompanied by the nominee’s curriculum vitae and statement of acceptance of nomination.

Biographical information for the following candidates will be enclosed with the ballots mailed to members in September.

Slate of Candidates

GENERAL ELECTION

President-Elect: James McCarthy, Harvard University; Richard Meserve, Carnegie Institution of Washington.

Board of Directors: Linda Katehi, University of Illinois, Urbana-Champagne; Clark Spencer Larsen, Ohio State University; Cherry Murray, Lawrence Livermore National Laboratories; David Tirrell, California Institute of Technology.

Committee on Nominations: Floyd Bloom, Neurome Inc.; Rita Colwell, University of Maryland, College Park; Thomas Everhart, California Institute of Technology; Mary Good, University of Arkansas, Little Rock; Jane Lubchenco, Oregon State University; Ronald Phillips, University of Minnesota; Robert Richardson, Cornell University; Warren Washington, National Center for Atmospheric Research.

SECTION ELECTIONS

Agriculture, Food, and Renewable Resources

Chair Elect: Roger N. Beachy, Washington University, St. Louis; Brian A. Larkins, University of Arizona, Tucson.

Member-at-Large of the Section Committee: Charles J. Arntzen, Arizona State University; James D.

Murray, University of California, Davis.

Electorate Nominating Committee: Douglas O. Adams, University of California, Davis; Richard A. Dixon, Samuel Roberts Noble Foundation; Sally A. Mackenzie, University of Nebraska, Lincoln; James E. Womack, Texas A&M University.

Anthropology

Chair Elect: Eugenie C. Scott, National Center for Science Education; Emõke J. E. Szathmáry, University of Manitoba.

Member-at-Large of the Section Committee: Leslie C. Aiello, Wenner-Gren Foundation for Anthropological Research; Dennis H. O'Rourke, University of Utah.

Electorate Nominating Committee: Daniel E. Brown, University of Hawaii, Hilo; Kathleen A. O'Connor, University of Washington; G. Phillip Rightmire, Binghamton University, SUNY; Payson Sheets, University of Colorado, Boulder.

Council Delegate: Michael A. Little, Binghamton University, SUNY; Ellen Messer, Brandeis University.

Astronomy

Chair Elect: Alan P. Boss, Carnegie Institution of Washington; Jill Cornell Tarter, SETI Institute.

Member-at-Large of the Section Committee: Carey Michael Lisse, Johns Hopkins University Applied Physics Laboratory; Tammy A. Smecker-Hane, University of California, Irvine.

Electorate Nominating Committee: Alan Marscher, Boston University; Heidi Newberg, Rensselaer Polytechnic Institute; Saeqa Dil Vrtilek, Smithsonian Astrophysical Observatory; Alwyn Wooten, National Radio Observatory.

Council Delegate: Guiseppina (Pepi) Fabbiano, Smithsonian Astrophysical Observatory; Heidi B. Hammel, Space Science Institute, Boulder.

Atmospheric and Hydrospheric Sciences

Chair Elect: Robert Harriss, Houston Advanced Research Center; Anne M. Thompson, Pennsylvania State University.

Member-at-Large of the Section Committee: Peter H. Gleick, Pacific Institute; James F. Kasting, Pennsylvania State University.

Electorate Nominating Committee: Walter F. Dabberdt, Vaisala, Inc.; Jennifer A. Francis, Rutgers University; Jack A. Kaye, Science Mission Directorate; Patricia Quinn, NOAA Pacific Marine Environmental Laboratory.

Biological Sciences

Chair Elect: H. Jane Brockmann, University of Florida; Mariana Wolfner, Cornell University.

Member-at-Large of the Section Committee: Anne L. Calof, University of California, Irvine; Yolanda P. Cruz, Oberlin College.

Electorate Nominating Committee: Kate Barald, University of Michigan; Joel Huberman, State University of New York, Buffalo; Maxine Linial, University of Washington; Jon Seger, University of Utah.

Council Delegate: Lois A. Abbott, University of Colorado, Boulder; Enoch Baldwin, University of California, Davis; Brenda Bass, University of Utah; Nancy Beckage, University of California, Riverside; Doug Cole, University of Idaho; Michael Cox, University of Wisconsin; Charles Ettensohn, Carnegie-Mellon; Toby Kellogg, University of Missouri; Catherine Krull, University of Michigan; J. Lawrence Marsh, University of California, Irvine; Michael Nachman, University of Arizona; David Queller, Rice University ; Laurel Raftery, Massachusetts General Hospital; Edmund Rucker, University of Missouri, Columbia; Johanna Schmitt, Brown University; Gerald B. Selzer, National Science Foundation; Diane Shakes, College of William and Mary; Rob Steele, University of California, Irvine.

Chemistry

Chair Elect: Steven L. Bernasek, Princeton University; Wayne L. Gladfelter, University of Minnesota.

Member-at-Large of the Section Committee: Dennis A. Dougherty, California Institute of Technology; Galen D. Stucky, University of California, Santa Barbara.

Electorate Nominating Committee: Gregory C. Fu, Massachusetts Institute of Technology; Joseph A. Gardella Jr., State University of New York, Buffalo; Linda C. Hsieh-Wilson, California Institute of Technology; Thomas Kodadek, University of Texas Southwestern Medical Center.

Council Delegate: Andreja Bakac, Iowa State University; Jon Clardy, Harvard Medical School; Mark A. Johnson, Yale University; C. Bradley Moore, Northwestern University; Buddy D. Ratner, University of Washington; Nicholas Winograd, Pennsylvania State University.

Dentistry and Oral Health Sciences

Chair Elect: Adele L. Boskey, Hospital for Special Surgery; Mary MacDougall, University of Alabama, Birmingham.

Member-at-Large of the Section Committee: Susan W. Herring, University of Washington; Paul H. Krebsbach, University of Michigan.

Electorate Nominating Committee: Luisa Ann DiPietro, University of Illinois, Chicago; Pete X. Ma, University of Michigan; Frank C. Nichols, University of Connecticut, Farmington; Ichiro Nishimura, University of California, Los Angeles.

Education

Chair Elect: George D. Nelson, Western Washington University; Gordon E. Uno, University of Oklahoma, Norman.

Member-at-Large of the Section Committee: Jay Labov, National Research Council; Gerald Wheeler, National Science Teachers Association.

Electorate Nominating Committee: Jeanette E. Brown, Hillsborough, NJ; Cathryn A. Manduca, Carleton College; Carlo Parravano, Merck Institute for Science Education; Jodi L. Wesemann, American Chemical Society.

Engineering

Chair Elect: Larry V. McIntire, Georgia Institute of Technology/Emory University; Priscilla P. Nelson, New Jersey Institute of Technology.

Member-at-Large of the Section Committee: Morton H. Friedman, Duke University Medical Center; Debbie A. Niemeier, University of California, Davis.

Electorate Nominating Committee: Mikhail A. Anisimov, University of Maryland, College Park; Rafael L. Bras, Massachusetts Institute of Technology; Melba M. Crawford, University of Texas, Austin; Corinne Lengsfeld, University of Denver.

General Interest in Science and Engineering

Chair Elect: Larry J. Anderson, Centers for Disease Control and Prevention; Barbara Gastel, Texas A&M University.

Member-at-Large of the Section Committee: Lynne Timpani Friedmann, Friedmann Communications; Renata Simone, WGHB Boston.

Electorate Nominating Committee: Earle M. Holland, Ohio State University; Don M. Jordan, University of South Carolina; Earnestine Psalmonds, National Science Foundation; Susan Pschorr, Platypus Technologies, LLC.

Geology and Geography

Chair Elect: Victor R. Baker, University of Arizona, Tucson; Richard A. Marston, Kansas State University.

Member-at-Large of the Section Committee: Sally P. Horn, University of Tennessee, Knoxville, Lonnie G. Thompson, Ohio State University.

Electorate Nominating Committee: Kelly A. Crews-Meyer, University of Texas, Austin; Sherilyn C. Fritz, University of Nebraska, Lincoln; Carol Harden, University of Tennessee; Neil D. Opdyke, University of Florida, Gainesville.

Council Delegate: William E. Easterling, Pennsylvania State University; Douglas J. Sherman, Texas A&M University.

History and Philosophy of Science

Chair Elect: Noretta Koetge, Indiana University; Thomas Nickels, University of Nevada, Reno.

Member-at-Large of the Section Committee: Karen A. Rader, Virginia Commonwealth University; Robert C. Richardson, University of Cincinnati.

Electorate Nominating Committee: Richard M. Burian, Virginia Polytechnic Institute and State University, Blacksburg; David C. Cassidy, Hofstra University; Mark A. Largent, Michigan State University; Kathryn M. Olesko, Georgetown University.

Industrial Science and Technology
Chair Elect: David L. Bodde, Clemson University; Stan Bull, National Renewable Energy Laboratory.
Member-at-Large of the Section Committee: Carol E. Kessler, Pacific Center for Global Security; Thomas Mason, Oak Ridge National Laboratory.
Electorate Nominating Committee: Ana Ivelisse Aviles, National Institute of Standards and Technology; Micah D. Lowenthal, The National Academies; Joyce A. Nettleton, Consultant, Denver, CO; Aaron Ormond, Global Food Technologies.

Information, Computing, and Communication
Chair Elect: Jose-Marie Griffiths, University of North Carolina, Chapel Hill; Michael R. Nelson, IBM Corporation.
Member-at-Large of the Section Committee: Christine L. Borgman, University of California, Los Angeles; Elliot R. Siegel, National Library of Medicine/NIH.
Electorate Nominating Committee: Gladys A. Cotter, U.S. Geological Survey; Deborah Estrin, University of California, Los Angeles; Richard K. Johnson, American University; Fred B. Schneider, Cornell University.

Linguistics and Language Science
Chair Elect: David W. Lightfoot, National Science Foundation; Frederick J. Newmeyer, University of Washington.
Member-at-Large of the Section Committee: Catherine N. Ball, MITRE Corporation; Wendy K. Wilkins, Michigan State University.
Electorate Nominating Committee: Miriam Butt, University of Konstanz; Barbara Lust, Cornell University; Robert E. Remez, Barnard College; Sarah G. Thomason, University of Michigan.

Mathematics
Chair Elect: William Jaco, Oklahoma State University; Warren Page, City University of New York.
Member-at-Large of the Section Committee: Jagdish Chandra, George Washington University; Claudia Neuhauser, University of Minnesota.
Electorate Nominating Committee: Frederick P. Greenleaf, New York University; Bernard R. McDonald, Arlington, VA; Juan Meza, Lawrence Berkeley National Laboratory; Francis Sullivan, Institute for Defense Analyses.
Council Delegate: James H. Curry, University of Colorado, Boulder; Joel L. Lebowitz, Rutgers University.

Medical Sciences
Chair Elect: Gail H. Cassell, Eli Lilly & Co.; Neal Nathanson, University of Pennsylvania Medical Center.
Member-at-Large of the Section Committee: Rafi Ahmed, Emory University, Atlanta; R. Alan B. Ezekowitz, Harvard Medical School.
Electorate Nominating Committee: Carl June, Abramson Family Cancer Research Institute; Michael Lederman, University Hospitals of Cleveland; Ronald Swanstrom, University of North Carolina, Chapel Hill; Peter F. Weller, Harvard Medical School.

Neuroscience
Chair Elect: John H. Byrne, University of Texas Medical School/Health Science Center, Houston; John F. Disterhoft, Northwestern University.
Member-at-Large of the Section Committee: Gail D. Burd, University of Arizona, Tucson; Charles D. Gilbert, Rockefeller University.
Electorate Nominating Committee: Theodore W. Berger, University of Southern California; György Buzsáki, Rutgers University; Alison Goate, Washington University School of Medicine, St. Louis; Gianluca Tosini, Morehouse University School of Medicine.
Council Delegate: Patricia K. Kuhl, University of Washington; Lynn C. Robertson, University of California, Berkeley.

Pharmaceutical Science
Chair Elect: Kenneth L. Audus, University of Kansas, Lawrence; Danny D. Shen, University of Washington.
Member-at-Large of the Section Committee: Michael Mayersohn, University of Arizona, Tucson; Ian A. Blair, University of Pennsylvania.
Electorate Nominating Committee: Charles N. Falany, University of Alabama, Birmingham; Kenneth W. Miller, American Association of Colleges of Pharmacy; John D. Schuetz, St. Jude Children's Research Hospital; Dhiren R. Thakker, University of North Carolina, Chapel Hill.

Physics
Chair Elect: Anthony M. Johnson, University of Maryland, Baltimore County; Cherry Murray, Lawrence Livermore National Laboratory.
Member-at-Large of the Section Committee: Sally Dawson, Brookhaven National Laboratory; Noémie B. Koller, Rutgers University.
Electorate Nominating Committee: Sanjay Banerjee, University of Texas, Austin; Elizabeth Beise, University of Maryland, College Park; Barbara Gross Levi, Stanford University; Pierre Meystre, University of Arizona, Tucson.
Council Delegate: Leonard J. Brillson, Ohio State University; W. Carl Lineberger, University of Colorado, Boulder, Luz J. Martínez-Miranda, University of Maryland, College Park; Miriam P. Sarachik, City College of New York.

Psychology
Chair Elect: Lila Gleitman, University of Pennsylvania; Randy Nelson, Ohio State University.
Member-at-Large of the Section Committee: Mike Fanselow, University of California, Los Angeles; Morton Gernsbacher, University of Wisconsin at Madison.
Electorate Nominating Committee: Richard Doty, University of Pennsylvania; Merrill Garrett, University of Arizona; John Kihlstrom, University of California, Berkeley; Martin Sarter, University of Michigan.

Social, Economic, and Political Sciences
Chair Elect: David L. Featherman, University of Michigan.
Member-at-Large of the Section Committee: Ronald J. Angel, University of Texas, Austin; Arnold Zellner, University of Chicago.
Electorate Nominating Committee: Gary L. Albrecht, University of Illinois at Chicago; Henry E. Brady, University of California, Berkeley; Gary King, Harvard University; Alvin E. Roth, Harvard University.

Societal Impacts of Science and Engineering
Chair Elect: Lewis M. Branscomb, University of California, San Diego; Eric M. Meslin, Indiana University.
Member-at-Large of the Section Committee: Ruth L. Fischbach, Columbia University; James Kenneth Mitchell, Rutgers University.
Electorate Nominating Committee: Ann Bostrom, Georgia Institute of Technology; Halina Szejnwald Brown, Clark University; Robert Cook-Deegan, Duke University; David B. Resnik, National Institute of Environmental Health Sciences/NIH.

Statistics
Chair Elect: William Butz, Population Reference Bureau; William Eddy, Carnegie-Mellon University.
Member-at-Large of the Section Committee: Robert E. Fay, Bureau of the Census; Francoise Seiller-Moiseiwitsch, Georgetown University Medical Center.
Electorate Nominating Committee: Norman Breslow, University of Washington; Marie Davidian, North Carolina State University; Fritz Scheuern, National Opinion Research Center; Judith Tanur, Stony Brook University.

Published by AAAS

Origins of HIV and the Evolution of Resistance to AIDS

Jonathan L. Heeney,[1]* Angus G. Dalgleish,[2] Robin A. Weiss[3]

The cross-species transmission of lentiviruses from African primates to humans has selected viral adaptations which have subsequently facilitated human-to-human transmission. HIV adapts not only by positive selection through mutation but also by recombination of segments of its genome in individuals who become multiply infected. Naturally infected nonhuman primates are relatively resistant to AIDS-like disease despite high plasma viral loads and sustained viral evolution. Further understanding of host resistance factors and the mechanisms of disease in natural primate hosts may provide insight into unexplored therapeutic avenues for the prevention of AIDS.

Human immunodeficiency viruses HIV-1 and HIV-2, the causes of AIDS, were introduced to humans during the 20th century and as such are relatively new pathogens. In Africa, many species of indigenous nonhuman primates are naturally infected with related lentiviruses, yet curiously, AIDS is not observed in these hosts. Molecular phylogeny studies reveal that HIV-1 evolved from a strain of simian immunodeficiency virus, SIVcpz, within a particular subspecies of the chimpanzee (*Pan troglodytes troglodytes*) on at least three separate occasions (*1*). HIV-2 originated in SIVsm of sooty mangabeys (*Cercocebus atys*), and its even more numerous cross-species transmission events have yielded HIV-2 groups A to H (*2*, *3*). The relatively few successful transfers, in contrast to the estimated >35 different species of African nonhuman primates that harbor lentivirus infections, indicate that humans must have been physically exposed to SIV from other primate species, such as African green monkeys. However, these SIV strains have not been able to establish themselves sufficiently to adapt and be readily transmitted between humans. Thus, it is important to understand the specific properties required for successful cross-species transmission and subsequent adaptation necessary for efficient spread within the new host population. Notably, among the three SIVcpz ancestors of HIV-1 that have successfully crossed to humans, only one has given rise to the global AIDS pandemic: HIV-1 group M with subtypes A to K. Here, we survey genetically determined barriers to primate lentivirus transmission and disease and how this has influenced the evolution of disease and disease resistance in humans.

Origins and Missing Links

A new study of SIVcpz not only confirms that HIV-1 arose from a particular subspecies of chimpanzee, *P. t. troglodytes*, but also suggests that HIV-1 groups M and N arose from geographically distinct chimpanzee populations in Cameroon. Keele *et al*. (*1*) combined painstaking field work collecting feces and urine from wild chimpanzee troupes with equally meticulous phylogenetic studies of individual animals and the SIV genotypes that some of them carry. These data have enabled a more precise origin of HIV-1 M and N to be determined. The origin of group O remains to be identified, but given the location of human cases, cross-species transmission may have occurred in neighboring Gabon.

Although HIV-1 has clearly come from SIVcpz, only some of the extant chimpanzee populations harbor SIVcpz. SIVcpz itself appears to be a recombinant virus derived from lentiviruses of the red capped mangabey (SIVrcm) and one or more of the greater spot-nosed monkey (SIVgsn) lineage or a closely related species (*4*). Independent data reveal that chimpanzees can readily become infected with a second, distantly related lentivirus (*5*), suggesting that recombination of monkey lentiviruses occurred within infected chimpanzees, giving rise to a common ancestor of today's variants of SIVcpz, which were subsequently transmitted to humans (Fig. 1A).

It is tempting to speculate that the chimeric origin of SIVcpz occurred in chimpanzees before subspeciation of *P. t. troglodytes* and *P. t. schweinfurthii*. However, this proposed scenario raises several questions: Why is SIVcpz not more widely distributed in all four of the proposed chimpanzee subspecies? Why is it so focal in the two subspecies in which it is currently found? These issues raise further questions regarding the chimpanzee's anthropology, its natural history, the modes of transmission of SIVcpz among chimpanzees, and the reasons that it is not a severe pathogen (*5*). These questions lead to other hypotheses that speculate about the intermediate hosts that might have given rise to SIVcpz and ultimately to HIV-1 (Fig. 1, B and C).

Diversity

Although the interspersal of SIVcpz and SIVsm in the molecular phylogeny of HIV-1 and HIV-2, respectively, reveals successful cross-species transmission events, there are a surprisingly limited number of documented cases, and direct evidence of a simian-to-human transmission is still missing. This suggests that, in contrast to a fulminant zoonotic (a pathogen regularly transmitted from animals to humans), a complex series of events (for instance, adaptations and acquisition of viral regulatory genes such as *vpu*, *vif*, *nef*, and *tat* and structural genes *gag* and *env*) was required for these SIVs to infect a human and to sustain infection at levels sufficient to become transmissible within the local human population. Closer examination of HIV-1 and HIV-2 groups and subgroups reveals differences in variants and genetic groups and rates of transmission in different populations even after infection is well established. This complex picture is beginning to merge with our understanding of the dynamics of evolving lentiviral variants that infect the natural nonhuman primate hosts. For instance, within the eight HIV-2 groups, A and B are endemic, whereas the others represent single infected persons clustering closely to SIVsm strains (*2*, *6*). These observations reinforce the notion that important adaptations have been necessary for the virus to acquire the ability to be efficiently transmitted.

Since its emergence, HIV-1 group M has diverged into numerous clades or subtypes (A to K) as well as circulating recombinant forms (CRFs) (*7*). There appears to have been an early ''starburst'' of HIV-1 variants leading to the different subtypes. CRFs have segments of the genome derived from more than one subtype, and two of these—CRF01_AE in Southeast Asia and CRF02_AG in West Africa—have relatively recently emerged as fast-spreading epidemic strains. Currently, subtype C and subtype A + CRF02_AG account for approximately 75% of the 14,000 estimated new infections that occur daily worldwide.

Regarding HIV in the Americas, subtype B was the first to appear in the United States and the Caribbean, heralding the epidemic when AIDS was first recognized in 1981. Subtype B remains the most prevalent (>80%) throughout the Americas, followed by undetermined CRFs (9%), F (8%), and C (1.5%) (*7*). There is a particularly high degree of genetic diversity of HIV-1 in Cuba, unparalleled in the Americas and similar to Central Africa (*8*), perhaps be-

[1] Department of Virology, Biomedical Primate Research Centre, Rijswijk 2280 GH, Netherlands. [2] St. George's Hospital Medical School, Division of Oncology, Department of Cellular and Molecular Medicine, Cranmer Terrace, London SW17 0RE, UK. [3] Wohl Virion Centre, Division of Infection and Immunity, University College, London W1T 4JF, UK.

*To whom correspondence should be addressed. E-mail: heeney@bprc.nl

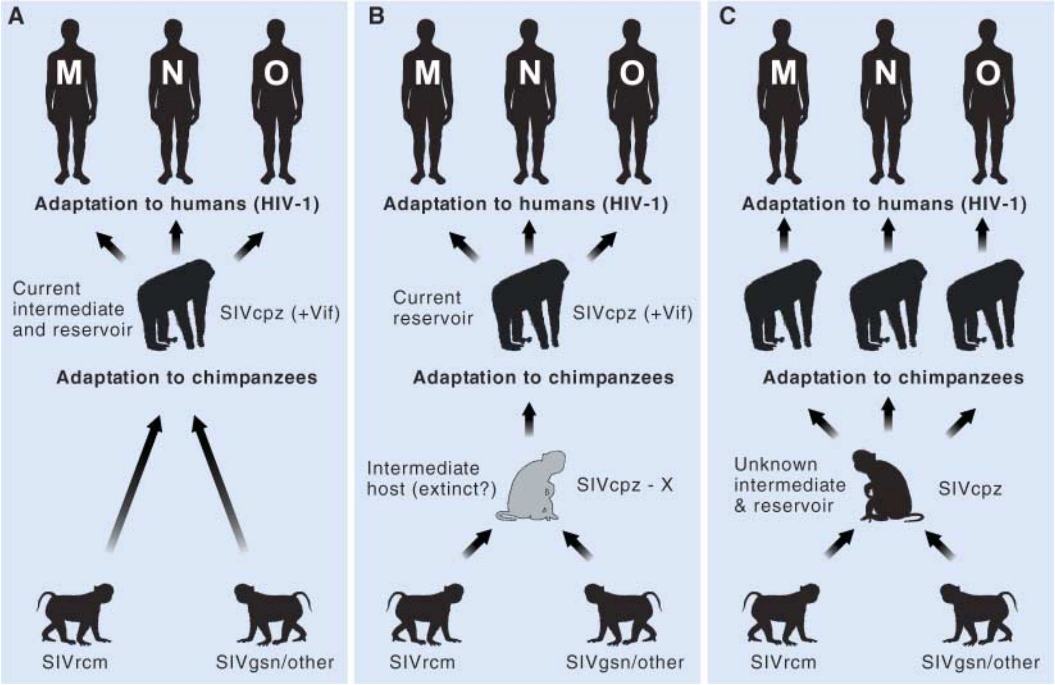

A
M
N
O
Adaptation to humans (HIV-1)
Current intermediate and reservoir
SIVcpz (+Vif)
Adaptation to chimpanzees
SIVrcm
SIVgsn/other
B
M
N
O
Adaptation to humans (HIV-1)
Current reservoir
SIVcpz (+Vif)
Adaptation to chimpanzees
Intermediate host (extinct?)
SIVcpz - X
SIVrcm
SIVgsn/other
C
M
N
O
Adaptation to humans (HIV-1)
Adaptation to chimpanzees
Unknown intermediate & reservoir
SIVcpz
SIVrcm
SIVgsn/other

Fig. 1. Possible cross-species transmission events giving rise to SIVcpz as a recombinant of different monkey-derived SIVs. Three different scenarios are considered. (**A**) *P. t. troglodytes* as the intermediate host. Recombination of two or more monkey-derived SIVs [likely SIVs from red capped mangabeys (rcm), and the greater spot-nosed (gsn) or related SIVs, and possibly a third lineage]. Recombination requires coinfection of an individual with one or more SIVs. Chimpanzees have not been found to be infected by these viruses. (**B**) Unidentified intermediate host. The SIVcpz recombinant develops and is maintained in a primate host that has yet to be identified, giving rise to the ancestor of the SIVcpz/HIV-1 lineage. *P. t. troglodytes* functions as a reservoir for human infection. (**C**) An intermediate host that has yet to be identified, which is the current reservoir of introductions of SIVcpz into current communities of *P. t. troglodytes* and *P. t. schweinfurthii*, as a potential source of limited foci of diverse SIVcpz variants.

cause Cuban troops served there for the United Nations. Less than 50% of Cuban infections are subtype B, and sequences of all subtypes are represented either as subtypes or in CRFs. The incidence of subtype C appears to be increasing rapidly in Brazil, just as it has in Africa and in East Asia.

Host-Pathogen Evolution

Upon adaptation of the virus to a new host, Darwinian selection would not only apply to the virus and host, but also to the modes of transmission between individuals in the new species, as well as to efficient replication within the infected individual (*9*). The modes of transmission of SIV likely differ from species to species. For example, parenteral transmission from bites and wounds as a consequence of aggression may be the main route of transmission in many nonhuman primates (*5*), whereas the major current mode of HIV transmission among humans is sexual. Nevertheless, parenteral transmission may well have played a more important role early in the emergence of the African epidemic (*10*), and it remains a risk today when nonsterile injecting equipment is used. Thus, efficient HIV transmission across mucosal surfaces may be a strongly selected secondary adaptation by the virus, given that humans tend to inflict minor parenteral injuries on each other less frequently then simians.

Whether genetic properties of the virus determine the rapid spread of HIV-1 subtypes such as C and CRF02_AG is not clear, although relative to other subtypes, subtype C appears to be present at higher load in the vaginas of infected women (*11*). It is not yet apparent whether certain subtypes are more virulent than others for progression to AIDS, although some indications of differences do exist (*12*).

SIVs do not appear to cause AIDS in their natural African hosts (Table 1). Similar to humans, however, several species of Asian macaques (*Macaca* spp.) develop AIDS when infected with a common nonpathogenic lentivirus of African sooty mangabeys (SIVsm became SIVmac). This observation demonstrates the pathogenic potential of such viruses after cross-species transmission from an asymptomatic infected species to a relatively unexposed naïve host species. Furthermore, SIV infection of macaques has provided a powerful experimental model system in which specific host as well as viral factors can be controlled and independently studied (*13*).

During the AIDS pandemic, it has become clear that host genetic differences between individuals as well as between species affect the susceptibility or resistance of disease progression, revealing a clinical spectrum of rapid, intermediate, or slow progression or, more rarely, nonprogression to AIDS within infected populations. A range of distinct genetic host factors, linked to the relative susceptibility or resistance to AIDS, influence disease progression. In addition to those genes that affect innate and adaptive immune responses, recently identified genes block or restrict retroviral infections in primates (including the human primate). These discoveries provide a new basis for detailed study of the evolutionary selection and species specificity of lentiviral pathogens.

Among the most important antiviral innate and adaptive immune responses of the host post-infection are those regulated by specific molecules of the major histocompatibility complex (MHC) (*13*). It is conceivable that in the absence of a vaccine or antiviral drugs, the human population will evolve and ultimately adapt to HIV infection, in much the same way that HIV is evolving and adapting to selective pressures within its host. Indeed, examples of similar host-viral adaptation and coevolution are evident in lentivirus infections of domestic animals. Nevertheless, greater insight into CD4 tropic lentiviruses and acquired resistance to AIDS has come from African nonhuman primates, which are not only reservoirs giving rise to the current human lentivirus epidemic but also possible reservoirs of past and future retroviral plagues.

Host Resistance Factors Influencing HIV Infection and Progression to AIDS

In humans, a spectrum of disease progression has emerged. Within the infected population, there are individuals with increased susceptibility as well as increased resistance to infection, who display rapid or slow progression to AIDS, respectively. Analyses of several large AIDS cohorts have revealed polymorphic variants in loci that affect virus entry and critical processes for the intracellular replication of lentivirions as well as subsequent early innate and especially highly specific adaptive host responses (*14*). To date, there is a growing list of more than 10 genes and more than 14 alleles that have a positive or negative effect on infection and disease progression (Table 2).

Polymorphic loci that limit HIV infection include the well-described *CCR5Δ32* variants

(*15*, *16*). The chemokine ligands for these receptors also influence disease progression: One example is Regulated on Activation Normal T Cell Expressed and Secreted (RANTES) (encoded by *CCL5*), with which elevated circulating levels have been associated with resistance to infections and disease. Moreover, it is the combination of polymorphisms controlling levels of expression of ligands and their specific receptors that exerts the most profound effect on HIV susceptibility and progression to AIDS; for example, gene dosage of *CCL3L1* acts together with *CCR5* promoter variants in human populations (*17*).

After retrovirus entry into target cells, intracellular "restriction factors" provide an additional barrier to viral replication. To date, three distinct antiviral defense mechanisms effective against lentiviruses have been identified: TRIM5α, a tripartite motif (TRIM) family protein (*18*); apolipoprotein B editing catalytic polypeptide (APOBEC3G), a member of the family of cytidine deaminases (*19*); and Lv-2 (*20*). TRIM5α restricts post-entry activities of the retroviral capsids in a dose-dependent manner (*18*, *21*), and the human form of this protein has apparently undergone multiple episodes of positive selection that predate the estimated origin of primate lentiviruses (*22*). The species-specific restriction of retroviruses is due to a specific SPRY domain in this host factor, which appears to have been selected by previous ancestral retroviral epidemics and their descendant endogenous retroviral vestiges. TRIM5α proteins from human and nonhuman primates are able to restrict further species of lentiviruses and gamma-retroviruses, revealing a host-specific effect on recently emerged lentiviruses.

The cytidine deaminase enzymes APOBEC3G and APOBEC3F also represent post-entry restriction factors that act at a later stage of reverse transcription than TRIM5α and are packaged into nascent virions. The APOBEC family in primates consists of nine cytosine deaminases (cystosine and uracil) and two others that possess in vivo editing functions (*19*, *23*). In the absence of the lentivirus accessory gene "*virion infectivity factor*" (*vif*), APOBEC3G becomes incorporated into nascent virions and inhibits HIV activity by causing hypermutations that are incompatible with further replication. At the same time, this represents a potentially risky strategy for the host, given that in some circumstances it might provide an opportunity for viral diversification (*24*). As with the primate TRIM5α family, APOBEC3G activity shows species-specific adaptations (*25*) emphasizing that coevolution of lentiviruses was a prerequisite for adaptation to a new host after cross-species transmission (*26*). Thus, although APOBEC3G clearly possessed an ancient role in defense against RNA viruses, a function that predates estimates of the emergence of today's primate lentiviruses, APOBEC3G appears to remain under strong positive selection by exposure to current RNA viral infections (*27*).

Evolving Host Resistance in the Face of New Lentiviral Pathogens

Failing the establishment of productive infection by the earliest innate defenses, natural killer (NK) cells of the immune system sense and destroy virus-infected cells and modulate the subsequent adaptive immune response. At the same time, the potentially harmful cytotoxic response of NK cells means that they are under tight regulation (*28*), which is centrally controlled by a raft of activating and inhibitory NK receptors and molecules encoded by genes of the MHC. Viruses have a long coevolutionary history with molecules of the immune system and a classical strategy for evading the cytotoxic T cell response of the adaptive immune system is by altering antigen presentation by MHC class I-A, I-B, or I-C molecules (*29*). In turn, the NK response has evolved to sense and detect viral infection by activities such as the down-regulation of class I MHC proteins.

Human lymphoid cells protect themselves from NK lysis by expression of the human MHC proteins human lymphocyte antigen (HLA)–C and HLA-E as well as by HLA-A and HLA-B. HIV-1, however, carries accessory genes, including *nef*, that act to differentially decrease the cell surface expression of HLA-A and HLA-B but not HLA-C or HLA-E (*30*). Such selective down-regulation may not only facilitate escape from cytotoxic T lymphocytes (CTLs) that detect antigens presented in the context of these MHC proteins but also escape from NK surveillance that might be activated by their loss of expression. However, within human MHC diversity, there may be an answer to the deception of NK cells by HIV. Certain alleles of HLA (HLA-Bw4) have been found to act as ligands for the NK inhibitory receptor (KIR) KIR3DSI and correlations with slower rates of progression to AIDS in individuals with the HLA-Bw4 ligand have been made with the corresponding expression of KIR3DSI expression on NK cells (*31*). The strength of this association between increased NK cell killing and HIV progression will have to bear the test of time as well as the test of the epidemic.

In the event that rapidly evolving pathogens such as HIV are able to evade innate defenses, adaptive defenses such as CTLs provide mechanisms for the recognition and lysis of new virus-infected targets within the host. This recognition depends on the highly polymorphic MHC class I molecules to bind and present viral peptides. However, a long-term CTL response will only be successful if the virus does not escape it through mutation. Additionally, it is advantageous to maintain MHC variability for controlling HIV replication and slowing disease progression (*32*), given that a greater number of viral peptides will be recognized if the infected individual is heterozygous for HLA antigens.

More importantly, there are qualitative differences in the ability of individual class I molecules to recognize and present viral peptides from highly conserved regions of the virus. These differences are observed in the spectrum of rapid, intermediate, and slow progressors in the HIV-infected human population (Table 2). Independent cohort studies have demonstrated the effects of specific HLA class I alleles on the rate of progression to AIDS with acceleration conferred by a subset of HLA-B*35 (HLA-B*3502, HLA-B*3503, and HLA-B*3504) specificities (*33*, *34*). Most notably, HLA-B*27 and HLA-B*57 have been associated with long-term survival. Both of these class I molecules restrict CTL responses to HIV by presenting peptides selected from highly conserved regions of Gag. Mutations that allow escape from these CTL-specific responses arise

Table 1. Natural lentivirus infections without immunopathology in African nonhuman primates.

Naturally resistant species and features of resistance
Examples
Chimpanzees (*P. troglodytes*), SIVcpz (HIV-1 in humans)
Sooty mangabeys (*C. atys*), SIVsm (HIV-2 in humans)
African green monkeys (AGMs) (*Chlorocebus* sp.), SIVagm
Common features of asymptomatic lifelong infection
Persistent plasma viremia
Maintenance of peripheral CD4 T cell levels
Sustained lymph node morphology
High mutation rate in vivo
Marginal increase in apoptosis returning to normal range
Transient low-level T cell activation and proliferation, returning to normal range
Less rigorous T cell responses than those in disease-susceptible species
Observed in one of these species, awaiting confirmation in others
High replication of virus in gastrointestinal tract, transient loss of CD4 T cells
CTL responses to conserved viral epitopes
Maintenance of dendritic cell function
Early induction of transforming growth factor–β1 and FoxP3 expression in AGMs with renewal of CD4 and increase in IL-10

only at great cost to viral fitness, reflected in lower viral loads (*13*) and survival benefit.

Evidence is emerging that HIV-1 is continuing to adapt under pressure from HLA-restricted immune responses in the human population. In a study that examined the relationship between HIV reverse transcriptase sequence polymorphisms and HLA genotypes, virus load was found to be predicted by the degree of HLA-associated selection of viral reverse transcriptase sequence (*35*). In a broader context, these results indicate that HLA alleles in the host population play an important role in shaping patterns of adaptation of viral sequences both within the host and at large.

Recent data have also started to suggest a potential influence that the HIV-1 epidemic may have on descendants of the HIV-infected population. In examining the relative contributions of HLA-A, HLA-B, and HLA-C alleles on restricting effective antiviral CTL, Kiepiela *et al*. (*36*) observed that HLA-B but not HLA-A allele expression influenced the rate of disease progression in that cohort. Thus, certain HLA-B alleles that favor long-term survival with HIV infection, in the absence of treatment, will be positively selected and will continue to evolve more rapidly over time. This coevolution of virus and host would be predicted to continue over generations until a relative equilibrium is reached between host resistance genes and virus infection. This would perhaps be similar to the asymptomatic lentivirus infections currently observed in naturally infected African nonhuman primates.

Disease Resistance in African Nonhuman Primates

Studies of SIVs in their natural hosts have been difficult and limited because of ethical issues and the endangered status of some species. For the most part, SIV natural history studies have been restricted to chimpanzees, sooty mangabeys, and African green monkeys. The chimpanzee is the closest living relative of humans, and two of its subspecies—*P. t. schweinfurthii* in East Africa and *P. t. troglodytes* in Western Central Africa—have certain wild communities with infected individuals (*1*). Although we should be cautious with generalizations, differences in transmission patterns may exist between the naturally infected monkey and ape populations (*5*). The prevalence of naturally occurring SIVsm in sooty mangabeys and SIVagm in African green monkeys appears to be relatively high, between 30 and 60%, increasing with age. However, SIVcpz infection across remaining free-ranging chimpanzee populations appears to be relatively low and regionally focal, restricted to certain troupes or communities in which it may reach levels greater than 20% (*1*, *37*, *38*).

Table 2. Human genes identified that influence HIV infection and disease.

Gene products	Allele(s)	Effect
	Barriers to retroviral infection	
TRIM5α	SPRY species specific	Infection resistance, capsid specific
ABOBEC3G	Polymorphisms	Infection resistance, hypermutation
	Influence on HIV-1 infection	
Coreceptor/ligand		
CCR5	Δ32 homozygous	↓ Infection
CCL2, CCL-7, CCL11 (MCP1, MCP3, eotaxin), H7		↑ Infection
Cytokine		
IL-10	5′A dominant	↓ Infection
	Influence on development of AIDS	
Coreceptor/ligand		
CCR5	Δ32 heterozygous	↓ Disease progression
CCR2	164 dominant	↓ Disease progression
CCL5 (RANTES)	ln1.1c dominant	↑ Disease progression
CCL3L1 (MIP1α)	Copy number	↓ Disease progression
DC-SIGN	Promoter variant	↓ Parenteral infection
Cytokine		
IL-10	5′A dominant	↑ Disease progression
IFN-γ	179T dominant	↑ Disease progression
Innate		
KIR3DS1 (with *HLA-Bw4*)	3DS1 epistatic	↓ Disease progression
Adaptive		
HLA-A, HLA-B, HLA-C	Homozygous	↑ Disease progression
*HLA-B*5802*, *HLA-B*18*	Codominant	↑ Disease progression
*HLA-B*35-Px*	Codominant	↑ Disease progression
*HLA-B*27*	Codominant	↓ Disease progression
*HLA-B*57*, *HLA-B*5801*	Codominant	↓ Disease progression

Few naturally infected chimpanzees have been available for study (*1*), and much of the knowledge of the immune responses to lentivirus in this species has come from animals infected with HIV-1 strains in the late 1980s and 1990s (*39*). In contrast to pathogenic HIV-1 infection of humans or SIVmac infection of rhesus macaques, the hallmarks of lentivirus infection in chimpanzees include the absence of overt CD4 T cell loss, a lack of generalized immune activation, and the preservation of secondary lymphoid structure, specifically with respect to MHC class II antigen presenting cells (APCs) in infected lymph nodes (*39*, *40*). In addition, there is little increase in apoptosis or anergy and no marked loss of interleukin (IL)-2–producing $CD4^+$ T cells after infection (Table 1) (*41*, *42*). These findings further underscore the importance of maintaining intact dendritic cell function and CD4 T cell interaction, which are symptoms of early immune dysfunction in infected AIDS-susceptible species (*40*).

Notably, $CD8^+$ CTLs in chimpanzees recognize highly conserved HIV-1 Gag epitopes, which correspond to almost identical epitopes presented by HLA-B*57 and HLA-B*27 alleles of humans with nonprogression or slow progression to AIDS (*43*). A phylogenetic analysis of MHC class I alleles in chimpanzees as compared with humans reveals an overall reduction of HLA-A, HLA-B, and HLA-C lineages in chimpanzees. Furthermore, comparative analysis of intron 2 sequences strongly supports marked reduction in the MHC class I repertoire, especially in the HLA-B locus (*44*). These data imply that chimpanzees may have experienced a selective sweep, possibly caused by a viral epidemic in the distant past. We could envision such a selective sweep of the modern day human population in the HIV-1 pandemic (in the absence of antiretroviral therapy), with a strong positive selection for HLA-B alleles beneficial for long-term survival (*36*).

It is becoming clearer that infected chimpanzees are relatively resistant to developing AIDS, not because they control virus load better than humans (*45*), but because they avoid the immunopathological events that affect the function of lymphoid tissue in humans and macaques that progress to AIDS. Thus, certain African nonhuman primates, such as chimpanzees, serve as natural lentivirus reservoirs and sustain lentivirus infection without the immunopathology (*40*, *42*) (Table 1). Mature CD4 T cells of chimpanzees are susceptible to SIVcpz or HIV-1 infection and cytopathology, but unlike macaques and humans, chimpanzees retain the renewal capacity to replace and sustain sufficient numbers of immunologically competent CD4 T cells to maintain immunological integrity (*39*).

How Will Humans Evolve in the Era of Medical Intervention?

New generations of more effective antiviral drug combinations are being developed, as are strategies to reduce virus load and facilitate restoration of CD4+ T cell numbers. The opportunity to convert an HIV-1 viremic patient into an aviremic individual by antiviral chemotherapy is an achievable clinical aim (*46*). Concern remains over the resident proviral population in long-lived lymphocytes and in APCs. Under antiretroviral treatment, aviremic CD4+ T cell tropic primate lentiviruses may also share features with the true "slow" replicating lentiviruses of ruminants. The prototypic lentiviruses of sheep and goats infect and persist in APCs such as dendritic and monotype/macrophage lineages without overt plasma viremia (*47*). Disease development is asymptomatic until late stages and is extremely protracted. Even in the absence of viremia and CD4 T cell loss, symptoms associated with chronic inflammation develop insidiously in diverse tissues resulting in a range of clinical conditions including encephalitis, pneumonia, and arthritis. It is important to consider that after solving the side effects of antiviral therapies such as lipodystrophy, HIV-infected aviremic humans might develop such classical lentivirus symptoms over a longer period of time.

Clearly, prophylactic strategies such as vaccines to prevent infection are the ultimate public health goals. Failing this, there is abundant evidence of previous retroviral epidemics embedded within the human genome. These suggest that there are further undisclosed antiretroviral defenses, which have coevolved and will continue to coevolve in human populations in response to retroviral insurgents.

References and Notes

1. B. F. Keele *et al.*, *Science* **313**, 523 (2006); published online 25 May 2006 (10.1126/science.1126531).
2. F. Damond *et al.*, *AIDS Res. Hum. Retroviruses* **20**, 666 (2004).
3. M. L. Santiago *et al.*, *J. Virol.* **79**, 12515 (2005).
4. E. Bailes *et al.*, *Science* **300**, 1713 (2003).
5. J. L. Heeney *et al.*, *J. Virol.* **80**, 7208 (2006).
6. F. Gao *et al.*, *J. Virol.* **68**, 7433 (1994).
7. S. Osmanov, C. Pattou, N. Walker, B. Schwardlander, J. Esparza, *J. Acquir. Immune Defic. Syndr.* **29**, 184 (2002).
8. M. T. Cuevas *et al.*, *AIDS* **16**, 1643 (2002).
9. P. Kellam, R. A. Weiss, *Cell* **124**, 695 (2006).
10. E. Drucker, P. G. Alcabes, P. A. Marx, *Lancet* **358**, 1989 (2001).
11. G. C. John-Stewart *et al.*, *J. Infect. Dis.* **192**, 492 (2005).
12. P. Kaleebu *et al.*, *J. Infect. Dis.* **185**, 1244 (2002).
13. P. J. Goulder, D. I. Watkins, *Nat. Rev. Immunol.* **4**, 630 (2004).
14. S. J. O'Brien, G. W. Nelson, *Nat. Genet.* **36**, 565 (2004).
15. W. A. Paxton *et al.*, *Nat. Med.* **2**, 412 (1996).
16. R. Liu *et al.*, *Cell* **86**, 367 (1996).
17. E. Gonzalez *et al.*, *Science* **307**, 1434 (2005).
18. M. Stremlau *et al.*, *Nature* **427**, 848 (2004).
19. A. M. Sheehy, N. C. Gaddis, J. D. Choi, M. H. Malim, *Nature* **418**, 646 (2002).
20. C. Schmitz *et al.*, *J. Virol.* **78**, 2006 (2004).
21. G. J. Towers, *Hum. Gene Ther.* **16**, 1125 (2005).
22. S. L. Sawyer, L. I. Wu, M. Emerman, H. S. Malik, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2832 (2005).
23. P. Turelli, D. Trono, *Science* **307**, 1061 (2005).
24. V. Simon *et al.*, *PLoS Pathog.* **1**, e6 (2005).
25. S. L. Sawyer, M. Emerman, H. S. Malik, *PLoS Biol.* **2**, E275 (2004).
26. H. P. Bogerd, B. P. Doehle, H. L. Wiegand, B. R. Cullen, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3770 (2004).
27. J. Zhang, D. M. Webb, *Hum. Mol. Genet.* **13**, 1785 (2004).
28. L. L. Lanier, *Annu. Rev. Immunol.* **23**, 225 (2005).
29. B. N. Lilley, H. L. Ploegh, *Immunol. Rev.* **207**, 126 (2005).
30. G. B. Cohen *et al.*, *Immunity* **10**, 661 (1999).
31. M. P. Martin *et al.*, *Nat. Genet.* **31**, 429 (2002).
32. M. Carrington, S. J. O'Brien, *Annu. Rev. Med.* **54**, 535 (2003).
33. M. Carrington *et al.*, *Science* **283**, 1748 (1999).
34. H. Hendel *et al.*, *J. Immunol.* **162**, 6942 (1999).
35. C. B. Moore *et al.*, *Science* **296**, 1439 (2002).
36. P. Kiepiela *et al.*, *Nature* **432**, 769 (2004).
37. E. Nerrienet *et al.*, *J. Virol.* **79**, 1312 (2005).
38. M. L. Santiago *et al.*, *J. Virol.* **77**, 7545 (2003).
39. J. L. Heeney, *Immunol. Today* **16**, 515 (1995).
40. E. Rutjens *et al.*, *Front. Biosci.* **8**, d1134 (2003).
41. M. L. Gougeon *et al.*, *J. Immunol.* **158**, 2964 (1997).
42. K. F. Copeland, J. L. Heeney, *Microbiol. Rev.* **60**, 722 (1996).
43. S. S. Balla-Jhagjhoorsingh *et al.*, *J. Immunol.* **162**, 2308 (1999).
44. N. G. de Groot *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11748 (2002).
45. P. ten Haaft *et al.*, *AIDS* **15**, 2085 (2001).
46. J. E. Gallant *et al.*, *N. Engl. J. Med.* **354**, 251 (2006).
47. S. Ryan, L. Tiley, I. McConnell, B. Blacklaws, *J. Virol.* **74**, 10096 (2000).

48. We thank T. de Koning and H. van Westbroek for assistance. This work was supported in part by grants from the NIH–Office of AIDS Research and NIH PO1 A148225-01A2 to J.L.H. R.A.W. is in part supported by the Medical Research Council. This work was in part facilitated by the Royal Society of Medicine.

10.1126/science.1123016

INTRODUCTION

Learning to Live With HIV

SINCE AIDS FIRST SURFACED IN LOS ANGELES IN 1981, INTERNATIONAL CONCERN HAS moved from the United States, Canada, and Europe to Africa and then to Asia. Now there's a growing appreciation that countries in Latin America and the Caribbean also have devastating HIV/AIDS epidemics, as well as some of the most creative and forceful responses seen anywhere. The following stories provide an in-depth look at both the epidemics and the responses, highlighting the affected communities, clinicians, researchers, and governmental and nongovernmental organizations alike.

Over the course of 9 months, *Science* correspondent Jon Cohen visited 12 countries that together represent the varied contours of the epidemic in this vast region, as well as the overlapping forces that drive HIV's spread. Cohen and photographer Malcolm Linton visited clinics, brothels, laboratories, shooting galleries, ministries of health, gay sex clubs, universities, slums, migrant way stations, prisons, and the homes of many people who struggle to live with the virus.

The Caribbean has been particularly hard hit, although the epidemic in Haiti appears to have peaked (p. 470). Heterosexual sex is the main mode of spread throughout the islands, and sex workers, some of whom cater to tourists (p. 474), often have high infection rates. Poverty and migration also fuel HIV's spread, as is apparent in the shantytowns that abut former sugar plantations in the Dominican Republic (p. 473). Puerto Rico has a staggering problem in injecting drug users (p. 475).

Throughout Mexico and Central America, men who have sex with men play a leading role in HIV's spread, although only the Mexican government has focused research and prevention campaigns on this population (p. 477). Honduras has novel programs to help descendents of African slaves known as Garifunas, who have a particularly high HIV prevalence (p. 481), and Belize is working to slow the spread among gang members (p. 483). Guatemala is struggling both to get a handle on the scale of its epidemic and to rapidly expand anti-HIV treatment to people most in need (p. 480).

Brazil dominates South America in its size, population, and the number of HIV-infected people who live there. The country has pioneered in offering "universal access" to antiretroviral treatment, but the escalating cost of the drugs poses a tremendous challenge (p. 484). In neighboring Argentina, the main mode of transmission has shifted from people injecting cocaine and men having sex with men to heterosexual sex (p. 487). And Peru, unlikely as it may seem, has become a research magnet for cutting-edge treatment and prevention trials (p. 488).

No countries in Latin America and the Caribbean have the double-digit prevalences frequently seen in sub-Saharan Africa, and its total population is not even half that of India, which alone has more infected people. Still, as should become clear at the XVI International AIDS Conference to be held from 13 to 18 August in Toronto, Canada, many opportunities exist to help countries in the region avoid some of the problems experienced elsewhere. And as these stories document, changes are desperately needed in many locales now, as HIV can be counted on to exploit every opportunity it can find.

A Kaiser Family Foundation Media Fellowship (www.kff.org) helped support Cohen's reporting for this project. All photographs are by Malcolm Linton. More of Linton's photographs, and the stories behind them, can be seen on *Science*'s Web site (www.sciencemag.org/sciext/aidsamericas).

–LESLIE ROBERTS AND JON COHEN

HIV/AIDS: Latin America & Caribbean

CONTENTS

News

See also related Editorial p. 409; Review p. 462; Report p. 523; and online material on p. 403 or at www.sciencemag.org/sciext/aidsamericas

CREDIT: MALCOLM LINTON

Published by AAAS

<< **New page.** Miriam Banks (*right*) was lucky to find quality care that remains inaccessible to many throughout the region.

OVERVIEW

The Overlooked Epidemic

As a Bible-toting evangelist moved from patient to patient and dispensed prayers in the women's AIDS ward at the Instituto Nacional del Tórax in Tegucigalpa, Honduras, Miriam Banks sat on her bed and flipped through an issue of *Vogue*. The magazine was stuffed with photos of impossibly glamorous models adorning stories about what to wear and where to shop. But on World AIDS Day on 1 December 2005, Banks, who had on hospital garb and a hairnet, was barely hanging on to her life. Banks, 24, lives on the island of Roatán, and her trip to the Honduran capital the month before required an airplane flight followed by a 7-hour bus ride, grueling even for the stout. Banks, who learned that she was infected with HIV 4 years earlier, arrived with tuberculosis, hypoplastic anemia, sinusitis, liver problems, and a CD4 cell count of just 33. (600 is the bottom end of normal.) But at the hospital, she had begun receiving anti-HIV drugs and was in a remarkably good mood. "The care is excellent here," she said in English, the main language of her island, to which she has since returned.

This aging hospital, one of Honduras's largest providers of HIV/AIDS care, provides a study in contrasts. So does the HIV/AIDS epidemic in Latin America and the Caribbean, which are home to diverse cultures, sexual mores, languages, patterns of drug use, ethnicities, and economic realities. "Living on the other side of the ocean, I used to look at the region as if it's all the same, but that's definitely not true," says epidemiologist Peter Piot, who heads the Joint United Nations Programme on HIV/AIDS (UNAIDS) in Geneva, Switzerland. "When it comes to AIDS, it's just not one place."

The epidemic in Latin America and the Caribbean has largely been overshadowed by the more severe problems in sub-Saharan Africa, the vastly larger population of Asia, and the attention that more developed countries have attracted with high-profile activism, substantial investments in finding solutions, and intense media coverage. But an estimated 2 million people live with HIV/AIDS in the region—more than the United States, Canada, Western Europe, Australia, and Japan combined. Half reside in the four largest countries: Brazil, Mexico, Colombia, and Argentina. Although far less populous, Haiti, the Bahamas, Guyana, Belize, and Trinidad and Tobago have the worst epidemics: Each has a prevalence above 2%. The virus is also moving from high-risk groups to the general population in Honduras, Guatemala, El Salvador, and Panama, where prevalences hover around 1%. "When I look at Latin America, I think Central America is the most vulnerable for the spread of HIV," says Piot.

Difficult as it is to assess the regional epidemic in Latin America and the Caribbean, HIV is aided and abetted by a few common factors: widespread poverty, massive migration, weak leadership, homophobia, tensions between church and state, and a dearth of research into patterns of transmission. Compounding the problems, HIV-infected people face pervasive stigma and discrimination, sometimes even from doctors and nurses.

As the epidemic varies, so have the responses of governments and nongovernmental organizations (NGOs). In many poor countries such as Honduras, it's difficult to find free antiretroviral drugs outside the major cities. But Haiti, which has the dual burden of being the poorest country in the region and the one with the highest HIV/AIDS prevalence, offers first-rate care in some very remote areas.

Although *machismo* leads many Latin American countries to play ostrich about homosexuality, Mexico and Peru each openly report that their epidemics are driven mainly by men who have sex with men (MSM)—including many who also have sex with women. The Caribbean, in contrast, largely has a heterosexual epidemic that's fueled by the popularity of sex workers, who do a thriving business with both locals and tourists. The church, a major cultural force throughout the region, has pressured politicians to block condom promotion in several countries. Yet in other areas, priests and nuns, working side by side with AIDS researchers and activists, run novel efforts to thwart the epidemic.

Changing course. Haiti's FOSREF teaches sex workers to become dance instructors.

Published by AAAS

The patterns of the epidemic continue to shift. Early on, for instance, injecting drug users (IDUs) played a prominent role in HIV's spread in the Southern Cone of South America; today IDUs are a major driver along the Mexico–U.S. border and in Puerto Rico and Bermuda. Meanwhile, massive migration both within the region and back and forth to the United States means that as the epidemic matures, the defining features of spread in each country begin to blur—as do the HIV strains that are circulating.

Subtype casting

Virologist Jean Carr of the Institute of Human Virology in Baltimore, Maryland, has worked with leading investigators throughout Latin America and the Caribbean to identify the subtypes of HIV spreading in different areas. "This tells you where the virus has been and where it's going," says Carr.

HIV-1, the main type of the virus responsible for the AIDS epidemic, now divides into nine subtypes. Evidence strongly suggests that subtype B first entered the Americas from Africa, likely coming to Haiti and then spreading to gay men in the United States, Canada, and Western Europe. In most countries of Latin America and the Caribbean, the epidemic emerged a few years later, again in gay men with subtype B, but the picture has since become much more complex.

Proper aim. Prevention programs work with men in this overcrowded Nicaraguan prison, but many countries ignore this and other high-risk populations.

In the Caribbean, Carr and her co-workers identified a distinctive form of subtype B—designated "B prime"—that has spread in Haiti, the Dominican Republic, Jamaica, and Trinidad and Tobago. Typically, she says, phylogenetic analyses cannot distinguish one subtype B from another. But on these Caribbean islands, B prime is distinct from the garden-variety B found elsewhere. And each of these islands has a predominantly heterosexual epidemic. "Is there a change the virus needs to do to become heterosexually transmitted, and is this phylogenetic analysis picking it up?" asks Carr.

The garden-variety B is the main subtype in Central and much of South America. But there is much more genetic diversity in the countries of the Southern Cone—southern Brazil, Paraguay, Uruguay, Argentina, and Chile. Subtype F, although not the major player, is prevalent in each of these countries. In Brazil, there's increasing spread of subtype C, too, which worldwide is the most common—and some researchers contend is also linked to heterosexual spread. Brazilian researchers have shown that this C most likely came from a single introduction from Africa.

Finally, around the globe HIV continues to increase its diversity by fusing subtypes together. Researchers have discovered several B/F recombinants, although only a few of these have spread much in Brazil, Argentina, and Uruguay. Carr notes that these B/F subtypes are mainly found in heterosexuals. "The bridge almost certainly is from IDUs and sex workers, not homosexuals," says Carr.

Mixed response

Across the region, increased political will, cheaper antiretroviral drugs, stronger NGOs, and the generous donations of bilateral and multilateral donors have combined to vastly improve access to treatment in recent years.

According to the World Health Organization (WHO), at the end of 2005, an estimated 315,000 people in Latin America and the Caribbean were receiving antiretroviral drugs. That's up from 210,000 people 2 years earlier, and it represents an impressive 68% coverage; worldwide, only 20% of the people most in need receive these drugs. "You have access to antiretrovirals in many, many places in Latin America and the Caribbean," says Brazilian epidemiologist Luiz Loures, who works with UNAIDS. "But it's a paradox. They are far behind when it comes to prevention for highly vulnerable populations like MSM and IDUs. My conclusion is it looks easier for a government to deal with treatment than prevention."

Throughout Latin America, MSM have significant epidemics, but in Central America and the Andean region of South America, in particular, tailored prevention efforts are few and far between. Transvestites, the group most discriminated against, have the highest prevalence of all—up to 45% in one Lima study—and receive the fewest services. A handful of countries have creative prevention programs for sex workers; the Haitian NGO FOSREF, for example, offers professional salsa lessons to women interested in leaving the business to become dance teachers themselves. But this population is often ignored, and female sex workers have double-digit prevalence in Central America, Suriname, Guyana, and on several Caribbean islands. Last in line to receive help in avoiding HIV are prisoners and IDUs, populations that frequently overlap and that are highly vulnerable to infection.

Tomorrow's challenge

Back at the Instituto Nacional del Tórax in Tegucigalpa, Elsa Palou, the head of infectious diseases, has witnessed firsthand the remarkable impact of potent antiretroviral drugs. Some 90% of treated patients, including Miriam Banks, responded to the therapy, and the treatment has decreased the annual mortality of AIDS cases from 43% to 9%. (Deaths mainly occurred in people who did not seek treatment until they had fewer than 50 CD4s.) But Palou is worried about the inevitable emergence of drug resistance and toxicities, "maybe in 5 years, maybe more, maybe less," she says. Brazil, which has treated more people with anti-HIV drugs for longer than any country in the region, already has seen a dramatic increase in the number of people who need to switch from their original drugs to more expensive regimens.

The total number of infected people will also likely continue to rise, although part of that climb is because potent drugs are allowing infected people to live longer. With the exception of Haiti, no country in Latin America or the Caribbean has seen a marked drop in HIV prevalence. By 2015, according to projections from WHO and UNAIDS, the 2 million HIV-infected people in Latin America and the Caribbean today will increase to nearly 3.5 million. Currently, AIDS claims 90,000 lives per year in the region. But between now and 2015, another 1.5 million Latin Americans and Caribbean islanders, at a minimum, are projected to die from the disease.

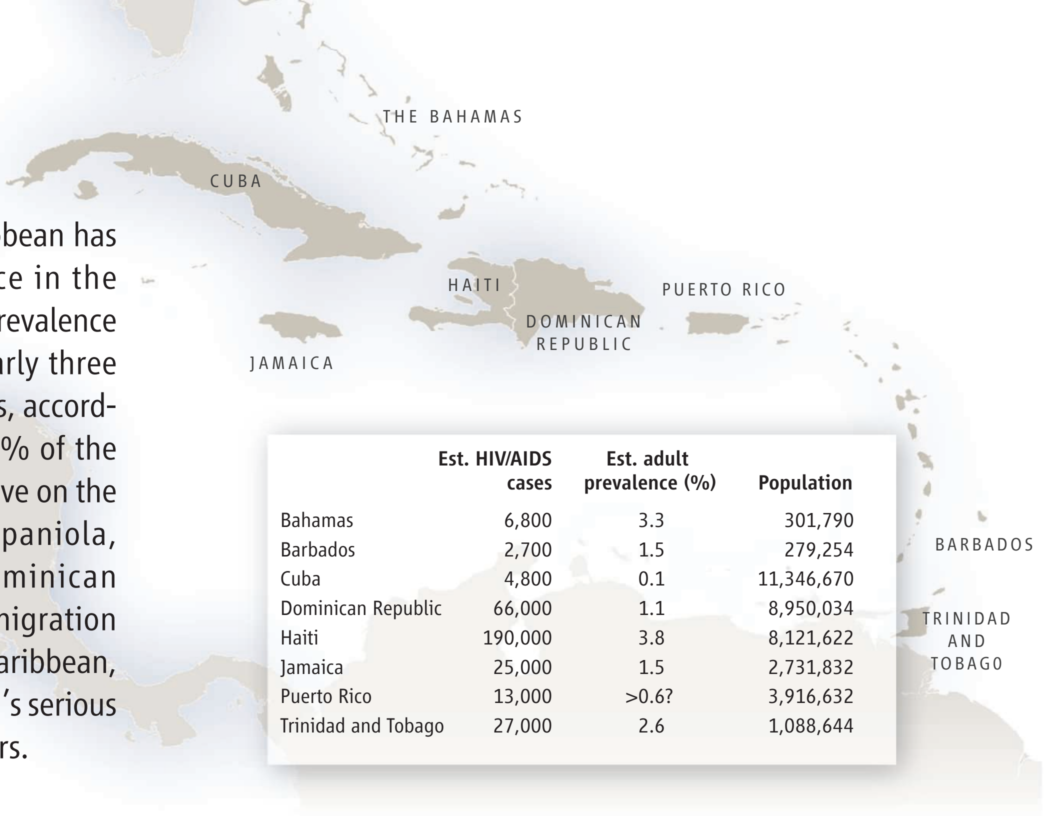
A surge in attention to HIV/AIDS may prove these projections wrong, and Latin America and the Caribbean will surely receive a boost in 2008 when Mexico becomes the first country in the region to host the massive International AIDS Conference. Then again, it's a tall order to contain the spread of HIV in any part of the world. But as the Spanish saying goes, *Con paciencia y saliva, el elefante se la metió a la hormiga*: With patience and saliva, the elephant can be put inside the ant.

–JON COHEN

Published by AAAS

The Caribbean

After sub-Saharan Africa, the Caribbean has the highest HIV/AIDS prevalence in the world. At the end of 2005, adult prevalence in the Caribbean was 1.6%—nearly three times higher than the United States, according to U.N. figures. More than 85% of the HIV-infected people in the region live on the heavily populated island of Hispaniola, home to both Haiti and the Dominican Republic. Heterosexual sex and migration drive the spread throughout the Caribbean, save for Puerto Rico's and Bermuda's serious HIV problems in injecting drug users.



	Est. HIV/AIDS cases	Est. adult prevalence (%)	Population
Bahamas	6,800	3.3	301,790
Barbados	2,700	1.5	279,254
Cuba	4,800	0.1	11,346,670
Dominican Republic	66,000	1.1	8,950,034
Haiti	190,000	3.8	8,121,622
Jamaica	25,000	1.5	2,731,832
Puerto Rico	13,000	>0.6?	3,916,632
Trinidad and Tobago	27,000	2.6	1,088,644

TABLE SOURCE: UNAIDS/CIA *WORLD FACT BOOK*

HAITI

Making Headway Under Hellacious Circumstances

This impoverished, conflict-ridden country is staging a feisty battle against HIV

PORT-AU-PRINCE, CANGE, AND CHAMBO, HAITI—Banners hang across the main thoroughfares in Port-au-Prince urging residents to report kidnappings. Blue-helmeted U.N. troops patrol the city in armored personnel carriers. The slums that border the once-elegant downtown have names like Cité Soliel and Bel Air that seem to mock their poverty and violence.

At an AIDS clinic called GHESKIO that sits at the edge of two of these slums, Cité L'Eternel and Cité de Dieu, the staff jokingly refers to the neighborhood as Kosovo. But the mood at GHESKIO (pronounced "jess-key-oh") is anything but hostile. The guards at the gates have no weapons, and as GHESKIO's founder and leader Jean "Bill" Pape likes to boast, "we have not lost one pencil" in the more than 20 years the clinic has operated there.

Pape climbs the stairs of the main clinic and enters the waiting room. About 100 patients, many spiffily dressed, sit in neat rows.

"*Bonjour*," says Pape.

"*Bonjour*!" the patients reply in unison.

Improbable as it seems, today *is* a good day for many of the people here, who receive antiretroviral drugs and state-of-the-art care they otherwise couldn't afford. It's also in many ways a good moment in the HIV/AIDS struggle in the country at large. The poorest country in the Western Hemisphere, Haiti has more HIV/AIDS patients per capita than any locale outside sub-Saharan Africa. Yet HIV-infected people here often receive better care than many in the Caribbean and Latin America, thanks largely to GHESKIO and another widely celebrated program, Zanmi Lasante—Creole for "Partners in Health"—started by medical anthropologist Paul Farmer of Harvard Medical School in Boston. And recently, encouraging signs have emerged that the epidemic in Haiti is shrinking.

Then again, combating HIV/AIDS in Haiti, where the ever-changing and crisis-plagued government has largely handed off its responsibilities to GHESKIO and Zanmi Lasante, remains an uphill battle. And it's a steep hill.

4H club

In 1982, a year after AIDS had first been diagnosed but not yet named in a cluster of homosexual American men in Los Angeles, the U.S. Centers for Disease Control and Prevention in Atlanta, Georgia, reported that a group of recent immigrants from Haiti had the strange opportunistic infections and immune problems that characterized the disease. Fears rose with reports of similar immune deficiencies among Haitians who still lived in that country. Soon, the mysterious ailment was being referred to as "the 4H disease," as it seemed to single out Haitians, homosexuals, hemophiliacs, and heroin users. "It was a disaster," says Pape, who at the time ran a rehydration clinic for children in conjunction with colleagues from Weill Medical College of Cornell University in New York City. "The tourism industry died. Nobody

Political outsider. GHESKIO's founder Jean "Bill" Pape strives to remain independent from the country's revolving door of political leaders. He says that has been a secret to GHESKIO's success.

Published by AAAS

wanted to come here. Even Haitians in the United States were afraid to come."

With help from Warren Johnson of Weill Cornell, Pape started GHESKIO (which stands for Groupe Haïtien d'Etude du Sarcome de Kaposi et des Infections Opportunistes). In 1983, Pape, Johnson, and co-workers published a landmark report in *The New England Journal of Medicine* (*NEJM*) that described how Haitians with AIDS had the same risk factors as Americans: men having sex with men, recipients of blood products, links to sex workers, and high rates of venereal diseases. Still, the notion that Haitians were somehow at a higher risk of contracting the disease persisted; theories flourished about links to voodoo or the predominance of swine flu. Worse yet, speculation surfaced that Haiti was responsible for the spread of AIDS to the United States. "There was all this prejudice against Haiti," says Pape, who still is visibly riled that epidemiologists pointed a finger at Haitians.

Although both Pape and Farmer have argued that HIV likely came to Haiti from the United States—gay men once flocked to the island as a tourist resort—molecular biological evidence suggests that HIV *did* arrive in Haiti earlier than anywhere else in the hemisphere. Further evidence connects the Haitian isolates to some found in Congo, a French-speaking country that recruited skilled Haitians after it gained independence in 1960. Two independent groups have published studies that date six early HIV isolates from Haitians to 1966–67, whereas the earliest non-Haitian samples in the United States trace back to the following year. "Both give the merest suggestion of Haiti being earlier—but with overlap in the error estimates," says Bette Korber, whose group at Los Alamos National Laboratory in New Mexico did one of the analyses.

Michael Worobey of the University of Arizona, Tucson, has recently recovered five "fossil" samples of HIV from Haitians diagnosed in the United States in the early 1980s that he says provide "absolutely crystal-clear evidence that the virus was in Haiti first." Worobey contends that understanding HIV's evolution may one day help vaccinemakers tailor preparations for specific regions. "All the B-subtype virus outside of Haiti comes from a single introduction that got into the homosexual population in the States and then Europe and went wild. And it required that raging wildfire to be seen."

Regardless of how HIV came to Haiti, the virus thrived, and by the end of 2001, the Joint United Nations Programme on HIV/AIDS (UNAIDS) estimated that 6.1% of the adults were infected. Studies by Pape and his co-workers in Haiti and at Weill Cornell have demonstrated that the vast majority of GHESKIO patients became infected through heterosexual sex. Disease progressed much more rapidly than in wealthy countries (7.4 years from infection to death versus 12 years), TB—which speeds HIV replication and thus immune destruction—was the most common AIDS-defining illness, and 6% of those coinfected with HIV and TB had dangerous, multidrug-resistant strains of the bacterium.

Testing 1, 2, 3. GHESKIO's well-equipped labs ran 35,000 HIV tests last year, as well as thousands of CD4 counts, TB stains, and rapid tests for syphilis.

By the end of 2005, reports UNAIDS, Haiti's adult prevalence had dropped to 3.8%. Pape contends that behavior change has led to this decline. Annual condom sales, he notes, jumped from less than 1 million in 1992 to more than 15 million a decade later. And GHESKIO studies show that sexually transmitted infections such as chancroid and genital ulcers, which can facilitate HIV transmission, have fallen steeply in their patients.

Analysis of these and other data conducted by Eric Gaillard of the Futures Group, a consulting firm funded by the U.S. government to help Haiti set HIV/AIDS policy, suggests that disease prevalence in the country has indeed dropped. But the researchers note that new infection rates—the incidence as opposed to the prevalence—started to decline about 15 years ago. This means that these behavior changes may have had less to do with the prevalence drop than other factors. "Overall, people died at a faster rate than others became infected," Gaillard and colleagues write in a paper in the April issue of *Sexually Transmitted Infections*. They also note that the prevalence drop coincides with the country's effort to prevent HIV transmission through blood transfusions (see graphs, p. 472).

Town and country

As a psychologist meets with rape victims in one of GHESKIO's cramped offices, lab techs in a nearby classroom watch a PowerPoint presentation about how HIV is transmitted. In another office, volunteers offering to join a trial of an experimental AIDS vaccine made by Merck take a test to make sure that their consent is truly informed. Technicians test samples of *Mycobacterium tuberculosis* for drug resistance in a lab outfitted with a special ventilation system. In another, sophisticated machines measure the level of the CD4 white blood cells that HIV preferentially targets and destroys. A long line of people, worried that they may have contracted HIV, syphilis, or another sexually transmitted infection, wait to have their blood drawn.

GHESKIO has slowly grown from a research-oriented AIDS clinic into something of an academic medical center that receives substantial funding from the U.S. National Institutes of Health. Pape ascribes part of GHESKIO's success to the fact that it's not part of the government. "If we were part of the Ministry of Health, we would have been dead," says Pape, explaining that it's had 24 ministers since 1986.

More than 3000 patients now receive anti-HIV drugs through GHESKIO. One of them is Elizabeth Dumay, a counselor and nurse assistant there. "Look at me," says an obviously robust Dumay, 42, who came to GHESKIO after losing both her husband and father to AIDS. At the time, her CD4 count was a mere 73 (600 to 1200 is normal). Today, Dumay has 603 CD4s, and virus levels in her blood are undetectable.

As the GHESKIO clinicians described in a December 2005 *NEJM* article, 90% of the 1000 AIDS patients they treated with potent antiretroviral drugs were alive after 1 year. Without the treatment, studies suggest that 70% of them would have died.

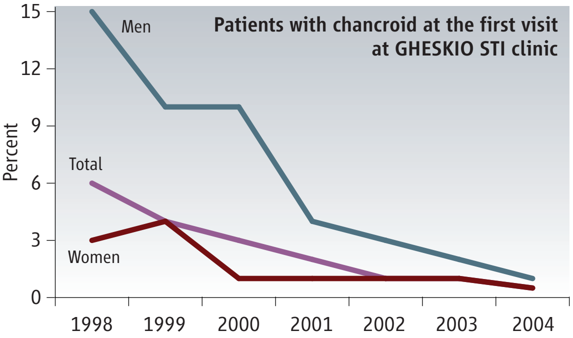
Pape has received a slew of accolades, including France's Legion of Honor. So has Farmer, who pioneered AIDS treatment in Haiti's rural Central Plateau. Farmer, who lives part-time in Haiti, is a MacArthur fellow, the subject of a popular biography, and the recipient of generous support from philanthropists. His group, Zanmi Lasante, now also has projects in Peru, Mexico, Guatemala, and Rwanda.

For more than 2 decades, Farmer has focused on improving health care in an impoverished part of the country that is only 56 kilometers from Port-au-Prince—but is a 3-hour journey by car on the rutted, mountainous roads. In 1998, Farmer launched an "HIV Equity Initiative" and began to treat poor, HIV-infected Haitians with antiretroviral drugs. When starting Zanmi Lasante, Farmer and his co-workers assailed the then-common wisdom that costs and lack of infrastructure made it impractical to use these medicines in poor countries. And, they wrote, if they can provide antiretroviral drugs "in the devastated Central Plateau of Haiti, it can be implemented anywhere."

Full house. Zanmi Lasante's inpatient ward in Cange doesn't have a bed to spare—and unfortunately can offer antiretroviral drugs only to the AIDS patients who live nearby.

Zanmi Lasante today has a sprawling medical campus in the rural town of Cange, which has been visited by the likes of Bill Gates Jr. (who flew in by helicopter). Farmer and his team of Haitian and Harvard doctors now provide antiretroviral treatment to 2000 patients at Cange and seven other sites. Zanmi Lasante also provides inpatient care, which GHESKIO doesn't. And, in an innovation borrowed from TB treatment, Zanmi Lasante assigns *accompagnateurs* to make home visits every day to observe patients taking their antiretroviral drugs. If doses aren't missed, they explain, HIV is less likely to develop resistance to the drugs.

Zanmi Lasante spent more than $10 million in Haiti last year, nearly twice GHESKIO's budget. Principal financial support for AIDS treatment for both groups comes from the Global Fund to Fight AIDS, Tuberculosis, and Malaria and the Bush Administration's President's Emergency Plan for AIDS Relief. Both GHESKIO and Zanmi Lasante also offer extensive training of health care workers, and they perform a combined 75,000 HIV tests each year.



Dramatic drop. At GHESKIO's clinic for sexually transmitted diseases, diagnoses of chancroid, which eases transmission of HIV, have steadily declined.

SOURCE (BOTTOM): JEAN W. PAPE/GHESKIO

Although their agendas overlap and they have much admiration for each other's work, Farmer and Pape have never published a paper together. "They have a research focus and we have a service focus," says Farmer, who has mainly written on issues of social justice and providing quality care in poor settings and whose group also offers comprehensive maternal care and builds new homes for people who live in shacks made of corrugated tin or wattle. "We're just using AIDS as our battle horse to get at poverty reduction. If we had the capacity to deliver the same quality of service we do now and do clinical trials, we would. One day, we're going to get there."

Meeting demand

Shortly before dawn on a March morning at the Zanmi Lasante campus, a few hundred people who have spent the night sleeping on the concrete benches and sidewalks that meander around the hilly grounds begin to rise. Some spent the night at this odd oasis—which features clinics, a hospital with two operating rooms, laboratories, training classrooms, a primary school, a church, and a warehouse filled with pharmaceuticals—because they saw a doctor too late in the day to return home; others wanted a good spot in line this morning. "We're being overwhelmed," says Farmer. "That's been the hardest part of our work."

At a new clinic that Zanmi Lasante recently opened about an hour's drive from Cange in Chambo, patients jam the waiting room all day for a chance to see one of two doctors on staff. Many of the patients are infected with HIV, but most have the same complaint: stomach pains. "I think it's just hunger," says Louise Ivers, a native of Ireland who treats HIV-infected people both in Haiti and at Massachusetts General Hospital in Boston. And her patients don't mince words. "I'm going to die if I don't get food to take with my medicine," complains an HIV-infected 24-year-old mother with three children in tow. A one-armed boy suddenly barges into the room unannounced. "The doctors told me to talk to you," says the boy, who explains that he lost his arm and his father in a car accident. Ivers refers him to the clinic's social worker. "It's very hard to know what to do," she says.

The inpatient hospital at Cange presents more wrenching dilemmas. The facility has several adults in the late stages of AIDS who are not eligible for anti-HIV drugs because Zanmi Lasante only offers antiretroviral drugs to people who live in areas where the group has *accompagnateurs*. "Until there's good care all across the country, we're going to get people coming from all over—and more from Port-au-Prince, ironically, than anywhere else," says Farmer. Last year, Zanmi Lasante's staff had 1.1 million visits with patients at clinics, and the *accompagnateurs* made 1.4 million more trips to patients' homes.

Although Zanmi Lasante has steadily won donor support and attracted local and foreign

Patient patients. Shortly after sunrise at Zanmi Lasante's Cange campus, long lines form at the clinic door.

doctors who want to work in rural Haiti, skeptics question whether the effort can be sustained. "Even if sustainability raises problems in 20 years, we didn't go in for a set timeline or to have projects with a beginning and an end," says Farmer. "We went in for the other 's' word: solidarity."

Increasing demand has burdened GHESKIO, too, which in October 2005 opened a second clinic in a less heavily trafficked part of Port-au-Prince. The pristine clinic abuts a vast, hardscrabble field, and a guard with a shotgun stands at its gate. "The neighbors don't know us here," shrugs Marie-Marcelle Deschamps, a clinician who helped Pape build GHESKIO. Already, the clinic is treating 400 HIV-infected people with antiretroviral drugs.

Despite all the progress, Pape estimates that at least 10,000 HIV-infected Haitians who need antiretroviral drugs immediately have yet to receive them. Still, like many other Haitians, he's hopeful that the election of René Préval in February will bring a measure of stability to the country—which should make it easier to combat HIV as vigorously as Pape, Farmer, and others would like. "You have to be an optimist here, despite all the odds," says Pape. "Otherwise, pack your bags and leave."

–JON COHEN

DOMINICAN REPUBLIC

A Sour Taste on the Sugar Plantations

Haiti's wealthier next-door neighbor is struggling to provide treatment to many HIV-infected people, and the problem's especially acute on the *bateyes*

SANTO DOMINGO, SAN PEDRO DE MACORÍS, MONTE PLATA, DOMINICAN REPUBLIC—The Dominican Republic shares the island of Hispaniola with Haiti, but the two countries could be across the globe from each other. Dominicans are Latin and pride themselves on their Spanish roots, whereas Haitians speak Creole and are largely descendents of freed African slaves. As tourists flock to the Dominican Republic each year, Haiti has seen its tourist industry evaporate over the past 2 decades. Dominicans have a vastly higher gross domestic product than their Haitian neighbors, whose average life expectancy is nearly 20 years shorter. And it follows that the two countries have starkly different HIV/AIDS epidemics that have attracted dramatically different responses. In an unusual twist, poorer and less stable Haiti is being celebrated for its pathbreaking AIDS efforts, largely led by two prominent nongovernmental organizations (NGOs). The Dominican Republic, on the other hand, is being lambasted for its shortcomings—the result, critics say, of government disinterest and outright obstructionism.

At the end of 2005, the Joint United Nations Programme on HIV/AIDS (UNAIDS) estimated that the virus had infected 1.1% of the adults in the Dominican Republic—a prevalence less than one-third of Haiti's. But according to insiders and outsiders alike, the Dominican Republic's HIV/AIDS programs in comparison are sorely lacking. "It's 1000 times better in Haiti," says Keith Joseph, a clinician at Columbia University who has done HIV/AIDS care in both countries. "It's astounding that a place with so much is unable to get things going."

Critical care. Sister Rivera provides *bateyes* with some medicines but does not have the anti-HIV or TB drugs that Miguel "Bebo" de Jesus needs.

Nowhere is this more evident than in the *bateyes*, where the Dominican epidemic is disproportionately concentrated. Originally built to house workers from Haiti on the sugar cane plantations, *bateyes* have become shantytowns largely filled with descendents of the original migrants or new Haitian immigrants. "People with AIDS in the *bateyes* are just dying without any kind of help," says Sister Concepcion Rivera, a nurse with the Sisters of Charity who runs a mobile health clinic.

The clinic attempts to care for people living in the many *bateyes* near San Pedro de Macorís, a port city on the southeast coast of the Dominican Republic. Although the van is stocked like a minipharmacy, Rivera, who has a master's degree in bioethics, on this March day has no anti-HIV drugs, nor can she treat tuberculosis, one of the biggest killers of people with AIDS. "On paper, the government does things, but in practice, they really provide nothing," says Rivera, adding that for the past 3 months the government has not even paid the small subsidy it promised her group.

Although the Dominican Republic now offers anti-HIV drugs in major cities such as Santo Domingo, Rivera's complaint repeatedly surfaces in the *bateyes*. Government studies showed that adult HIV prevalence was 5% in the *bateyes* in 2002 and jumped as high as 12% in men between 40 and 44 years old. And even where antiretroviral drugs are available, the government has faced intense criticism for moving slowly. UNAIDS estimates that 17,000 Dominicans need anti-HIV drugs, but as of December 2005, only 2500 received them through public programs.

Published by AAAS

The Sun. The Sand. The Sex.

BOCA CHICA, DOMINICAN REPUBLIC—At the Plaza Isla Bonita bar that stretches from the main downtown street to the beach, the cocktail waitresses dress in campy "Ship's Ahoy" outfits with sailor hats and midriff tops. When not serving high-octane rum drinks, they dance suggestively to the blaring merengue, bachata, and reggaeton music. Tables and bar stools fill with young Dominican women, who flirt aggressively with American, Dutch, German, and Italian men twice if not three times their age. Sanky Pankies—local young men who favor dreadlocks, bling bling, and tank tops—cruise the perimeter looking for foreign women or men.

The waitresses sing along when a popular song comes on by the band Mambo Violento: *Sin gorrito, no hay cumpleaño*—without a little hat, there is no birthday party. But in this case, a little hat is a condom, and the birthday party doesn't involve cake.

Sex tourism is booming in several of the resorts here, says Antonio de Moya, an epidemiologist and anthropologist who has long studied the subculture and works with the presidential AIDS program COPRESIDA. In the past 15 years, the Dominican Republic has become a tourist magnet, attracting 3.4 million vacationers in 2004, more than double the number who visited in 1991, according to the Caribbean Tourist Organization. And the Caribbean as a whole entertained more than 21 million tourists in 2004. Today, sex tourism and HIV/AIDS have become hot topics in Jamaica, Cuba, Barbados, the Bahamas, St. Lucia, St. Marteens, and Curaçao.

Sails job. The cocktail waitresses at the Plaza Isla Bonita bar attract male tourists, who often then find a sex worker offering her—or his—services.

Deanna Kerrigan, an international health specialist at the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland, studies sex work in the Dominican Republic. She stresses that outside resorts such as Boca Chica, tourists are not the main clients. "There is a very large local sex-work industry," says Kerrigan. Sex is sold everywhere, from brothels and rendezvous homes called *casas de citas* to discos and car washes. HIV prevalence in the country's estimated 100,000 female sex workers ranges from 2.5% to 12.4%, depending on the locale. Kerrigan says the places with lower prevalence reflect "intensive interventions" by nongovernmental organizations such as the one she collaborates with called the Centro de Orientación e Investigación Integral.

Sex workers of course could have both local and foreign clients, but three women working the main street here this warm winter evening insist that they avoid Dominicans. "A Dominican will pay 300 pesos and be on top of you for 2 hours," says Aracelis, as the other women laugh and nod their heads. "And they don't want to use condoms." Aracelis and her friends insist that *sin gorrito, no hay cumpleaño*, and all say they are HIV-negative. But they still worry. "The first thing I say when I leave the house in the morning is 'Please, God, take care of me,' " says Aracelis. Then, as though her prayers were answered, she notices an elderly German man. "He's my boyfriend, not a client," she says, prancing over to him. "He sends me money every month."

–J.C.

Still, NGOs have made some headway in both prevention and treatment programs. Family Health International (FHI), which is funded by the U.S. government, supports several of these programs, but its director in Santo Domingo, Judith Timyan, laments that this is necessary. "This country's relatively rich and has a huge middle class," says Timyan, who has since left to do HIV/AIDS work in Haiti. "The Dominican Republic should have grown out of its need for help."

Bad blood

In 1821, Haiti invaded the Dominican Republic and ruled for 22 years, creating bad blood that has yet to disappear. "The Dominican ruling class will tell you everything that's going wrong with the country is the fault of Haiti," says Geo Ripley, an ethnographer and artist who is a consultant on *bateyes* to the United Nations.

This bad blood in part explains the government's limited response to the problem in the *bateyes* and also discourages any attempt to replicate Haiti's HIV/AIDS successes. "If you say to the Dominican people, 'We can learn from Haiti,' they'd say, 'We don't have anything to learn from them,' " says Eddy Perez-Then, a clinician who is now completing a Ph.D. dissertation about *bateyes* near the southwestern city of Barahona.

As in Haiti, the Dominican epidemic initially involved men who have sex with men, but it has gradually become more "feminized" and driven by heterosexual sex. This is reflected in the ratio of men with AIDS to women, which in 1986 was 3.63:1 and today is nearing 1:1. Government researchers estimate that 78% of infections now occur through heterosexual sex, some of which is linked to a booming sex trade (see sidebar, at left): Some sex-worker communities have had documented prevalence above 12%.

Cultural mores regarding promiscuity may partly explain why the *bateyes* and Haiti have similarly high prevalences, but many experts suggest that's too simplistic a view. Nicomedes "Pepe" Castro, who has worked with *bateyes* for 28 years, notes that in the last century the sugar industry primarily attracted male migrants. "*Bateyes* were the only part of the country where the proportion of men was higher than women: 4 to 1." This, in turn, created more sharing of partners and a greater market for sex workers. With the demise of the sugar cane industry, Antonio de Moya, an epidemiologist and anthropologist who works with COPRESIDA—the presidential commission on AIDS—says an increasing number of young Haitians who immigrate are becoming sex workers themselves. Finally, and perhaps most important, the rampant poverty in the *bateyes* facilitates HIV's spread, which is tied to a lack of education and less access to prevention tools such as condoms and treatment of other sexually transmitted diseases.

Epidemiologist William Duke, who works with FHI, says it's unclear whether the Dominican epidemic is growing, shrinking, or stabilizing. "In general, our surveillance is very weak in the public health sector," says Duke. "When you go outside of the capital, it's difficult to catch the data." Although Haiti's surveillance surely has gaps, NGOs, government-run prenatal clinics,

Published by AAAS

and outside consultants have reliably tracked that epidemic.

Whereas Haiti in 2002 marshaled the strong support of then–First Lady Mildred Aristide and became one of the first countries to secure a grant from the Global Fund to Fight AIDS, Tuberculosis, and Malaria to buy anti-HIV drugs, the Dominican Republic did not make a similar deal until 2004. Haiti exceeded its targets for delivering antiretroviral drugs to people in need; the Dominican Republic, in contrast, has repeatedly lowered its sights.

Even today, one NGO in Santo Domingo, the Instituto Dominicano de Estudios Virologicos, provides care for 20% of the people receiving anti-HIV drugs. Ellen Koenig, an American clinician who has lived in the country since 1969 and started the institute, assails the attitude of the government that recently left office. "There were more people in the country living *from* AIDS than *with* AIDS," charges Koenig. "It was ridiculous."

Perez-Then says about 25% of the *bateyes* do have government clinics nearby, but the residents don't use them much. "They're afraid to go," he says. In some cases, they are recent Haitian immigrants who only speak Creole. Others do not have proper documentation or fear discrimination.

Perez-Then worries, too, about the complexity of treating HIV-infected people and the quality of care available at government-run programs. The Dominican Republic has one of the highest rates of drug-resistant tuberculosis in the world, which occurs when people start treatment but then miss doses of their pills. The same could easily happen with antiretroviral drugs, he says.

Taking it home

Weeds and scrub brush have overgrown the old sugar cane fields near Batey Cinco Casas, located in Monte Plata province a few hours' drive from Santo Domingo. But there's some new growth that has thrilled the residents: a clinic built by the Batey Relief Alliance. Similarly, the Christian relief group World Vision has built a clinic in Batey 6 near Barahona. Both clinics have a limited ability to help HIV-infected people, but they do what they can. In March, for instance, the Batey Relief Alliance was regularly transporting 28 HIV-infected people from the Monte Plata area to Santo Domingo to receive anti-HIV drugs. Many more need transportation, says Maria Virtudes Berroa, who runs the relief association's Santo Domingo office, but the organization doesn't have enough money. One of those is an emaciated man they recently found dying from late-stage AIDS. Like hundreds of thousands of Haitians before him, Jean-Claude Delinua, 31, moved to the Dominican Republic 11 years ago to cut cane. Delinua now lives on the edge of a fallow sugar cane plantation in a one-room shack. He rarely leaves his hammock, which is made from a pig-feed sack. He has no job, no family, no possessions beyond the clothes he wears, toiletries, a paperback, and a photograph of himself 8 months earlier when he was buff and hale. Delinua, who speaks in Creole, says he knows about the care offered in his home village in Haiti's Central Plateau. "I'd like to go back," says Delinua. "But I don't have the money, and I'm not sure my family would receive me."

Graham Greene, author of the classic novel about Haiti called *The Comedians*, once wrote that it was impossible to exaggerate the country's poverty. For HIV-infected people like Jean-Claude Delinua, it's all too easy to exaggerate the prosperity of the Dominican Republic.

–JON COHEN

PUERTO RICO

Rich Port, Poor Port

Good HIV/AIDS care and strong research in this U.S. commonwealth often mean little to the island's many heroin addicts

SAN JUAN, PUERTO RICO—If Viviana Valentin lived on any other Caribbean island, she'd likely be dead by now. Diagnosed with an HIV infection in 1990, Valentin has developed resistance to several antiretroviral drugs and once had a CD4 count of zero, an indicator that HIV had decimated her immune system. She has two children and no job. Yet today, Valentin is receiving T-20, the most expensive anti-HIV drug, which retails for more than $20,000 a year and requires twice-daily injections. She's also benefiting from state-of-the-art care at the University of Puerto Rico (UPR), where she is enrolled in a clinical trial studying neurological complications of the disease. "I have the best doctors," says Valentin, who was born and raised in New York City and moved to Puerto Rico when she was 21. "They've done a wonderful job."

As a commonwealth of the United States, Puerto Rico enjoys one of the strongest economies in the Caribbean, which supports not only the top-notch care many HIV-infected people receive but also a burgeoning research community. But that's the rosy picture. There are thorns as well. Puerto Rico's per capita income is lower than that of any state on the mainland. Because it is a U.S. territory, HIV/AIDS prevalence figures are lumped with those on the mainland, a practice that many experts think masks the extent of Puerto Rico's epidemic. "We're submerged into the U.S. statistics," says virologist Edmundo Kraiselburd, who directs both UPR's NeuroAIDS research program and the Caribbean Primate Research Center.

And unlike the epidemics in the rest of the Caribbean, Puerto Rico's is driven primarily by

Prickly issues. Injecting drug users at this San Juan shooting gallery have severely limited access to health care and drug substitutes such as methadone.

Published by AAAS

Ample Monkeys and Money Nurture Robust Research

SAN JUAN AND CAYO SANTIAGO, PUERTO RICO—This country's close ties to the United States, combined with its large colony of rhesus macaques of Indian origin, have spawned several collaborations with leading AIDS researchers from the mainland—a rarity in much of the Caribbean.

Rhesus macaques are the main model used to test AIDS vaccines, but they're in short supply. Cayo Santiago, a 15-hectare island off Puerto Rico that has been home to Indian macaques since 1938, has a surplus and must cull about 120 animals each year. Over the past 4 years, Edmundo Kraiselburd of the University of Puerto Rico estimates that UPR has shipped some 600 monkeys to various U.S. researchers, most of them studying AIDS. Some of these monkeys have also now been moved to the UPR campus, where Puerto Rican investigators, in collaboration with a group led by Thomas Folks of the U.S. Centers for Disease Control and Prevention in Atlanta, Georgia, are conducting AIDS vaccine studies.

Monkey business. UPR's Edmundo Kraiselburd runs a primate center and is helping to build an internationally recognized HIV/AIDS research community.

Kraiselburd also heads the NeuroAIDS Program, which teams Puerto Rican clinicians and basic researchers with neuroAIDS specialists on the mainland. The project, which began in 2001 with a $6 million grant from the U.S. National Institutes of Health (NIH), has several novel studies under way. One, led by Carlos Luciano, is comparing HIV-infected children and adults to try to unravel the link between HIV and peripheral neuropathy, the most common nerve complication of AIDS. In a separate study, neurologist Valerie Wojna and immunologist Loyda Meléndez are using proteomics to investigate the causes of HIV dementia.

With NIH support, Puerto Rican researchers have long participated in clinical trials of AIDS drugs. For instance, UPR's Carmen Zorrilla was a co-investigator of the landmark multisite study that in 1994 first proved that antiretroviral drugs could prevent HIV transmission from mother to infant. (UPR's medical center has had only one case of mother-to-child transmission since.) And recently, again with NIH backing, Puerto Rico joined the HIV Vaccine Trials Network and, separately, started an HIV/AIDS research collaboration among the country's three medical schools. Zorrilla, who is helping to lead both projects, is particularly excited about bringing together young researchers from institutions that have long competed with one another. "This is a small island," says Zorrilla. "These young investigators will inherit this AIDS problem, and they need to find the solutions."

–J.C.

injecting drug users (IDUs), who are often discriminated against at clinics or emergency rooms. "The doctors don't want them," says José "Chaco" Vargas Vidot, a clinician who in 1990 started an outreach program for IDUs called Iniciativa Comunitaria. Vargas Vidot complains that the country has too few methadone treatment clinics and needle-exchange programs, which elsewhere have proven key to lowering transmission rates. "The government is ignoring our AIDS epidemic," he charges.

So although Puerto Rico is indeed a rich port for patients such as Viviana Valentin and many HIV/AIDS researchers, IDUs often have a starkly different vantage.

Heroin hub

On an early weekday afternoon in a barrio outside San Juan called La Colectora, a dozen men and one woman pay $1 each to enter a shooting gallery, a small house where users inject and then typically collapse into a chair. Out front, two outreach workers and a doctor from Iniciativa Comunitaria set up a needle-exchange program. Julio, a 33-year-old heroin addict, shuffles up and lays eight syringes on the ground, receiving an equal number in exchange. Julio, who is homeless, does not shuffle because he is high: Injecting has left him with bloody and blackened abscesses on his calves that may be gangrenous, says Angel González, a clinician with the program.

Julio says the stench coming from his legs makes a bad situation even worse. He couldn't make it to his methadone treatment program, he says, because "they started to refuse to let me on the bus. … The smell was bad, and people would complain." He says an emergency room also sent him away without care.

González says Julio is one of many addicts the system has failed. "Patients have to go through so many obstacles to get treatments," says González. "We need big changes here." UPR's Carmen Albizu-García, who is conducting a small drug-substitution program with addicted prisoners, is also deeply frustrated by the official resistance to proven HIV prevention methods. "In Puerto Rico, we've been very, very hesitant to do what we have to do to control the epidemic," she says.

Heroin's popularity on the island has many roots, but it's clearly tied to its strategic location for South American traffickers. The Puerto Rican Department of Health says that half of the AIDS cases reported to date are heterosexual IDUs, while another 7% are IDU males who have sex with men. UPR obstetrician/gynecologist Carmen Zorrilla says that roughly two-thirds of 2000 HIV-infected women she is following were infected by having sex with men who were IDUs. The HIV/IDU situation in Puerto Rico is "a public health emergency," says Sherry Deren, director of the Center for Drug Use and HIV Research in New York City.

Deren, along with sociologist Rafaela Robles and epidemiologist Héctor Colón of the Central University of the Caribbean in Bayamón, Puerto Rico, led a provocative study comparing 399 IDUs in San Juan to 800 Puerto Rican IDUs living in New York City. Between 1996 and 2004, the researchers found, users in Puerto Rico injected nearly twice as frequently, favored mixtures of heroin and cocaine known as speedballs, and were more than three times as likely to share needles. Between 20% and 25% of the IDUs were infected in both locales, but the new infection rate in Puerto Rico (3.4% per year) was nearly four times higher. The study also found significantly fewer needle-exchange and methadone programs in Puerto Rico, and twice as many HIV-infected participants in New York were receiving antiretroviral drugs. Not surprisingly, the mortality rate in Puerto Rico was almost three times higher. If a city or state on the mainland had these statistics, says Deren, "I think there'd be much more attention given to the problem." Colón points a finger at policymakers who "still believe that treating drug users is a waste of money."

–JON COHEN

Published by AAAS

Mexico & Central America

HIV/AIDS relentlessly exploits the gaps that still separate the haves from the have-nots in this region. Free antiretroviral treatment is widely available, but it's often hard to find the drugs outside of major cities. Without money, it's even harder to find quality care. Epidemiological data suggest that men who have sex with men, rampant migration, a thriving sex-worker industry, gangs, and crowded prisons are all contributing to the spread of HIV. Honduras and Belize are the hardest hit; Nicaragua and Mexico are at the other end of the spectrum.

	Est. HIV/AIDS cases	Est. adult prevalence (%)	Population
Belize	3,700	2.5	279,457
Costa Rica	7,400	0.3	4,016,173
El Salvador	34,000	0.9	6,704,932
Guatemala	53,000	0.9	14,655,189
Honduras	59,000	1.5	6,975,204
Mexico	170,000	0.3	106,202,903
Nicaragua	5,800	0.2	5,465,100
Panama	15,000	0.9	3,039,150

TABLE SOURCE: UNAIDS/CIA WORLD FACT BOOK

MEXICO

Land of Extremes: Prevention and Care Range From Bold to Bleak

With a population more than twice as large as all of Central America combined, the country has the most HIV/AIDS cases in the region yet a relatively low prevalence

MEXICO CITY AND TIJUANA, MEXICO—In 2003, when the Mexican government appointed Jorge Saavedra to head CENSIDA, its top AIDS agency, the messages were unmistakable. Saavedra, an articulate spokesperson, is an openly gay and HIV-infected clinician in a country where—as in much of Latin America—an abundance of machismo causes serious cases of homophobia. He's also a prime example of the power of modern anti-HIV drugs. "He was dying from AIDS," says sociologist Mario Bronfman, a former top health official who hired Saavedra at the Ministry of Health years ago when no good anti-HIV drugs existed. "It's very symbolic that he's the head," says Bronfman, who now works with the Ford Foundation in Mexico City. "And not just because he's HIV-positive and gay. No one can understand the problem from the inside the way that Jorge can."

The choice of Saavedra was surprising even to those doing AIDS clinical care and research. "I could not believe that they chose him," says Luis Soto-Ramírez, one of Mexico's leading HIV/AIDS researchers, who welcomed the move. "It was amazing." But it's not the only unusual aspect of Mexico's epidemic—or the country's response to it.

In contrast to other countries in Latin America and the Caribbean, which tend to downplay the extent of the spread of HIV among men, Mexico candidly reports that the primary driver of its epidemic is men who have sex with men—many of whom do not consider themselves gay or bisexual. Since 2003, the government has also had a policy of universal access to antiretroviral drugs, and this year the government reported that everyone who has been identified with advanced disease is receiving treatment. In another sign of the country's progressiveness, activists, sex workers, and researchers have organized innovative efforts to combat the spread of HIV, as has Saavedra, who last year launched a provocative antihomophobia campaign.

Although Mexico has made big strides in tackling HIV/AIDS, there are still some glaring gaps, says Carlos del Rio of Emory University in Atlanta, Georgia, who headed AIDS policy for the Mexican government from 1992 to 1996. The epidemic has not grown as much as he and others once feared it would, but del Rio says the heterosexual spread in rural communities "is much more difficult to control." Research is often

Innovative approach. As part of its HIV prevention efforts for sex workers, the NGO Aproase discreetly oversees the transactions on Sullivan Boulevard all night long.

Published by AAAS

“primitive,” he says—in particular, prevalence data are thin—and collaborations remain rare. And although antiretroviral drugs may be widely available, many people who need them do not know they are infected, and pharmacies often run out of drugs. The training of clinicians, and thus the quality of care, is also spotty, del Rio says: “The lofty goal of universal access is not being fully realized.”

Prevalence puzzles

If you believe the official figure—and many experts don’t—only 0.3% of the adults in Mexico are infected with HIV. That’s half the U.S. prevalence. “It’s very difficult to say what’s happening in Mexico,” says Soto-Ramírez, who runs an HIV/AIDS lab and clinic at the National Institute of Nutrition in Mexico City. “The numbers say very different things from what *I* think.” From his vantage point, the prevalence must be higher—and increasing. “I’m seeing many more women and many more rural cases,” he says.

Epidemiologist Carlos Magis-Rodríguez, CENSIDA’s research director, has found a surprising degree of heterosexual spread in rural Mexican communities and disturbing new evidence that migration is a major factor. “We find a lot of at-risk behavior in these little towns,” says Magis-Rodríguez. In collaboration with the University of California’s Universitywide AIDS Research Program (UARP), Magis-Rodríguez’s team is comparing 1500 people from five Mexican states who in the past year migrated to California for seasonal work to some 1200 who did not. Preliminary data suggest that the migrants have more sexual partners, use drugs and alcohol more frequently, and hire sex workers more often.

A second study suggests that migrants are becoming infected in California and bringing the virus back to rural communities in Mexico at high rates. The researchers compared the prevalence of HIV in 800 Mexican migrants temporarily living in California (0.6%) to 1500 who migrated and then returned home to Mexico (1.1%). “Is it possible that a low-prevalence country like Mexico could take off like India and China?” asks epidemiologist George Lemp, who heads UARP in Oakland, California. “That’s of great concern.”

A separate collaboration between clinicians at Tijuana General Hospital (TGH) and researchers at the University of California, San Diego (UCSD), published in the January *Journal of Acquired Immune Deficiency Syndromes*, suggests that the prevalence among pregnant women—generally considered an indicator of spread in the population at large—may also be significantly higher than official estimates. CENSIDA reported in 1997 that only 0.09% of pregnant women in Mexico were infected with HIV. In the new work, UCSD’s Rolando Viani and co-workers tested more than 2500 pregnant women at TGH in 2003 who were either receiving prenatal care or who came to the hospital for the first time during labor. The group receiving prenatal care had a prevalence of 0.33%—nearly four times higher than earlier estimates. And in the group that only showed up in labor, which reported more frequent use of injecting drugs and more sexual partners, prevalence jumped to 1.12%.

Gynecologist Jorge Ruiz-Calderon, a coauthor at TGH, says the initial reaction to the study from colleagues and officials alike was anger and denial. “They wanted to cut our heads off,” he says. “Most of my colleagues don’t want to know anything about the problem.” Many critics also viewed TGH, which Ruiz-Calderon notes sees “the poorest of the poor” in a border town that attracts people from other locales, as an aberration. “They see these pregnant women as outcasts,” says Viani. And he says that’s a serious mistake: “Eventually,” he predicts, “miniepidemics like this one will interchange with the general population.”

Quality of care

Although TGH may not represent Mexico at large, it does illustrate the serious limitations that exist even in middle-income countries that have universal-access policies. Anti-HIV drugs can dramatically lower a pregnant woman’s risk of transmitting the virus to her baby. But at TGH—a well-equipped hospital in a large city that likely offers a higher standard of care than many other facilities in Mexico—screening of pregnant women is far from routine. Ruiz-Calderon says the residents and nurses are “not offering HIV tests to every pregnant woman, or they’re doing it after delivery.”

<< On the move. Tapachula’s Casa del Migrante provides temporary shelter—and HIV prevention education—to 7000 migrants each year.

Prevention Programs Target Migrants

TECÚN UMÁN, GUATEMALA, AND TAPACHULA, MEXICO—In late November 2005, more than a month after Hurricane Stan walloped Guatemala and southern Mexico, the border in Tecún Umán was still closed because of damage to the bridge that connects the two countries. But the unofficial border crossing remained open for business. From daybreak until sundown, rafts fashioned from truck tires and wood planks shuttled people across the Suchiate River that separates this spicy border town from Mexico. A policeman stood watch much of the time, gladly ignoring the illegal migration for a small fee.

HIV negotiates the border with similar ease, carried by the constant flow of people. And this border in particular has helped clarify the theory that migration is a significant driver of the AIDS epidemics in this region—and the world at large. “In the beginning, it wasn’t easy to convey the message that migration has something to do with HIV/AIDS,” says sociologist Mario Bronfman, an Argentinean native who in the 1990s led groundbreaking studies that looked at migrants in Tecún Umán and 10 other “transit stations” in Central America and Mexico. Bronfman, who works with the Ford Foundation in Mexico City, says, “Now that we have hard data, it’s very clear there is a problem.”

Bronfman’s studies assessed knowledge and opinions about HIV/AIDS at each transit station. As Bronfman and his colleagues reported in the journal *AIDS* in 2002, a long list of factors puts migrants at higher risk of HIV infection: poverty, violence, few available health services, increased risk-taking, rape, loneliness, and large numbers of sex workers—all of which aptly characterize Tecún Umán today. They also found women to be more vulnerable because of “transactional” and “survival” sex that they had in exchange for food or protection during their travels.

Educavida, a nongovernmental organization sponsored by the United Nations Population Fund to do HIV/AIDS education and prevention, targets the wide array of migrants who temporarily call this town home. “Some stop here because they’re thinking of the American dream, and this is a place along the route,” says Educavida’s director, psychologist Brigida Garcia.

MACHISMO
¡Tú Puedes Cambiarlo!
Tienen derecho a ser respetados
LA DISCRIMINACION
Sólo una cosa puede detenerlos...
Si no puedes ejercer tus derechos ¡DENUNCIA!

Confronting homophobia. CENSIDA head Jorge Saavedra launched a provocative campaign against discrimination against men who have sex with men.

Viani notes that UCSD has not had a case of mother-to-child transmission of HIV since 1994; TGH documented seven infected babies last year alone. TGH also routinely runs out of pediatric formulations of the anti-HIV drugs used to treat infected children. "We're 20 minutes away from San Diego, but things are so different," says coauthor Patricia Hubbard, who coordinates the binational research program.

To Nuar Luna, a prominent AIDS activist, the biggest challenge Mexico faces is unequal access to quality care. "If you have influence and you have money, you have access," says Luna, who has struggled to find competent care for his own HIV infection. "This is Mexico—and this is Latin America. It's a region with a lot of racism and classism and social issues. You can hear Jorge Saavedra say, 'Here in Mexico, we have full access.' But we have to analyze what kind of access we have. The good services are for the rich ones, and the bad services are for the poor."

Reaching out

Despite the many concerns that people at the front have about Mexico's response to HIV/AIDS, nongovernmental organizations (NGOs) and the government itself have launched several innovative prevention efforts. One takes place each evening in a Mexico City "dark room," a club where men meet to have sex. The HIV-prevention service offered by the NGO Ave de México gives new meaning to the word outreach.

Not only do workers from Ave de México pass out condoms and lubricants, but they also put their hands between men in flagrante delicto to make sure that they're using protection. Dentist Carlos García de León, who in his off hours runs the organization, says their studies found that nearly half of the men were not using condoms. "Most people accept it very well and are thankful," says García de León. "They say, 'I wasn't thinking.' " He notes that in a gay sex club in, say, the United States, this type of intervention wouldn't fly. "They'd kill you," he laughs.

Late at night on the city's Sullivan Boulevard, Alejandra Gil and her group Aprose offer another uniquely Mexican approach to prevention. Gil, a former sex worker, provides a comprehensive program to protect the women who line the street and try to catch the eyes of men driving by. In addition to providing counseling and a clinic that offers testing for sexually transmitted infections such as HIV, Gil and her adult son sit in cars all night long and oversee each transaction, transporting the women to nearby hotels for their rendezvous—and even going to the room if they take longer than usual. "If the women don't have security, we can't help them with their health issues," says Gil.

Another creative project has stepped up prevention efforts for injecting drug users in Tijuana, two-thirds of whom report never having been tested for HIV. A mobile health clinic travels around the city to areas that health care workers typically avoid, providing tests, clean syringes, and limited treatment. Delivering care at shooting galleries "takes away the stigma" that often prevents users from seeking help, says UCSD epidemiologist Steffanie Strathdee, who is running the project with Remedios Lozada, an AIDS clinician in Tijuana.

On the national front, Saavedra has spearheaded an antihomophobia campaign of radio and TV ads—so provocative that two Mexican states refused to run them—and posters, including one that shows a man and a woman both leaning their heads against the archetypical macho Mexican man dressed in revolutionary garb. "The antihomophobia campaign really has opened a lot of discussion on this issue," Saavedra says.

Saavedra agrees that the country has a long way to go in its prevention efforts. And he also concedes that the government's quick launch of a universal access program meant that many health care workers and clinics were not as well trained in using the drugs as he would have liked. "We needed to do that first step in order to stop a lot of people from dying," says Saavedra. "But I understand the way people feel and what they need. I'm part of them."

–JON COHEN

(No solid figures exist on how many Mexicans and Central Americans migrate to the United States each year, but experts estimate that they number more than 1 million.) Today's clients include a Nicaraguan mother of three who sells sex in one of the town's many brothel/bars, an Ecuadorian man en route to the United States, and an HIV-infected woman who was a U.S. resident for 12 years and returned to her hometown a few years ago. Educavida does HIV testing, but Hugo Rivera, a clinician who works with the group, says he has little to offer people who test positive other than a referral to other locales that have antiretroviral drugs. "You do the examinations, and then they leave," says Rivera.

And migration shows no sign of abating. Annelise Hirschmann, head of Guatemala's National AIDS Program, says the country's long-standing civil war that ended in 1996 still spurs migration, as families try to reunite. "The secondary issues that surround the war definitely feed the epidemic," she says. Studies have shown that Mayans, who constitute about half of the country's population, are also at high risk because they travel frequently for agricultural work. And Hurricane Stan is just the latest natural disaster to drive Guatemalans from their homes. "There's a mass exodus of young people going to the States right now because of Hurricane Stan," says Dee Smith, a Maryknoll sister in Coatepeque who runs the HIV/AIDS-oriented Proyecto Vida. "They had few opportunities *before* Stan."

No visa necessary. Migrants freely cross the Suchiate River between Guatemala and Mexico.

At the Casa del Migrante in Tapachula, Mexico—the closest big city and the first stop for many who cross at Tecún Umán—there is more hard evidence that migrants face an increased risk for HIV infection. This church-run lodging, which offers HIV/AIDS education, distributes a questionnaire to the 7000 people who pass through each year about their sexual lives during the journey. In 2004, fewer than 20% of the men reported having used condoms, and about 8% of the women said they had been raped. "Amigo Migrante," reads a poster near the entrance. "For HIV/AIDS, no border exists."

–J.C.

GUATEMALA

Struggling to Deliver on Promises And Assess HIV's Spread

Epidemiological data are scarce, and outside of the capital, so are antiretroviral drugs

COATEPEQUE, QUETZALTENANGO, AND GUATEMALA CITY, GUATEMALA—Over the past 7 years, Luz Imelda Lucas, 31, has become entirely too intimate with despair. First, HIV took the life of her husband, who she says also infected her. His parents were certain she had become infected first. "They told me I killed him and that I was going to die and my children were going to die," says Lucas, who lives in the southwestern town of Coatepeque. Lucas's youngest child died when she was 28 months old, she says. In 2002, Lucas's own days seemed numbered as her immune system bottomed out.

Then, in a stroke of great fortune, Médecins Sans Frontières (MSF) launched a new program in Coatepeque that offered free anti-HIV drugs. Lucas was selected as one of the first nine people in town to receive the medicines, and her health rebounded. Maryknoll sisters, Catholic missionaries who work in many countries, also hired her at their Proyecto Vida, which offers HIV/AIDS testing, counseling, and health care for infected people. Lucas officially is a nutritionist but is also something of a counselor. "I like to make it clear to people that having the virus, you can still be productive and continue living," says Lucas, who has a new boyfriend, too.

By the end of 2005, some 5500 HIV-infected people in Guatemala were receiving antiretroviral drugs, says Annelise Hirschmann, director of the country's National AIDS Program. Five years earlier, the only people being treated were the wealthy minority who could buy their own drugs, the small percentage protected by the country's social security system, and the few who enrolled in clinical trials. Roughly half of the drugs today come from MSF; the rest are purchased by the government or through a $40 million, 5-year grant awarded to the country in October 2004 by the Global Fund to Fight AIDS, Tuberculosis, and Malaria. Hirschmann says many people who were once selling their homes and preparing to die are now looking for jobs. But she acknowledges that there are far too many people who either don't know they are infected or have no access to the drugs, and "there are a lot of people dying from AIDS." Many sharply criticize the government for this because it passed a law in 2000 that said all Guatemalans had the right to treatment.

Arduous commutes. Eduardo Arathoon says the centralization of HIV/AIDS care at clinics like his in Guatemala City is badly hurting many Mayans who live in remote areas and must travel long distances.

One obstacle is that outside Guatemala City, free drugs are available at relatively few centers. "Most everything is centralized in this city," complains Eduardo Arathoon, who runs the Luis Angel García family clinic at Hospital San Juan de Dios in the capital. Arathoon points to an HIV-infected couple with their little girl. "The couple gets up at 3 a.m. and takes three buses to get here," he says. The centralization particularly hurts Mayans, who make up about half the population and often live in remote areas.

These problems will soon be compounded: MSF is leaving the country, which has Lucas and many other patients worrying about their futures once again.

Guesstimations

The Joint United Nations Programme on HIV/AIDS estimated at the end of 2005 that Guatemala had 71,000 HIV-infected people and an adult prevalence of 0.9%. But as in the rest of Central American, a dearth of surveillance makes it hard to get a good fix on the extent of the HIV/AIDS epidemic there—and thus how best to target prevention efforts. "Epidemiology is not seen as that important," says César Núñez, an epidemiologist based in Guatemala City who led the only in-depth studies of HIV's spread in Guatemala and other countries for the Central American HIV/AIDS Prevention Project (PASCA). "Countries and ministries of health are concerned that they have treatment for people in these countries. But we can't forget prevention either."

Funded mostly by the U.S. Agency for International Development, PASCA worked in 2001 and 2002 with the Guatemalan health ministry to measure HIV prevalence in high-risk groups. In men who have sex with men, the study found a prevalence of 11.5%. Nearly half of those men considered themselves bisexual or heterosexual rather than gay, putting their female partners at high risk, too. Female sex workers overall had a relatively low prevalence of 4.5%, but that figure jumped to 14.9% in women who worked the streets rather than in brothels, discos, or other "fixed" establishments.

PASCA had hoped that Guatemala and other countries would continue and expand the studies. "We were not an epidemiological surveillance system; we're the spark," says Núñez. But, says Edgar Monterroso, who heads the Guatemala City office of the U.S. Centers for Disease Control and Prevention (CDC), "none of the countries was able to pick up and do their own surveillance." CDC is now attempting to help Guatemala do these studies.

In particular, no one has properly evaluated HIV's spread among the Mayans, says Monterroso. But a small study conducted at the Luis Angel García clinic suggests that incidence may be three times higher in

Published by AAAS

Worried. Luz Imelda Lucas fears that she'll lose access to the anti-HIV drugs that have saved her life.

Mayans, who are often treated as second-class citizens, than in *ladinos*. "We think that group's more vulnerable," says Arathoon. Not only do many Mayans have trouble with Spanish, complicating prevention efforts, but they also have less access to health care in general. "We think that's where the epidemic will move," says Arathoon.

A study of patients at the government-run Rodolfo Robles tuberculosis hospital in Quetzaltenango supports that assertion. Between 1995 and 2002, HIV prevalence in TB patients at the hospital—74% of whom were Mayan—jumped from 4.2% to 12%. As of May 2005, no antiretroviral drugs were available in Quetzaltenango, the country's second-largest city.

Tough transitions

No one knows how many people are dying because they do not have access to antiretroviral drugs, says the National AIDS Program's Hirschmann. And even some of those taking the drugs are concerned about their continued supply because MSF announced in July 2005 that it was phasing out its program in Coatepeque, which now treats 500 people. Lucas is worried that the government will not respond adequately, and some Guatemalan AIDS clinicians and government AIDS officials share those concerns. "MSF obviously did something really good because they brought treatment to a country that wasn't offering it," says Hirschmann. "But they have created somewhat of a panic in patients on treatment. … I would be very afraid if I were a patient living with HIV and had to cross over to receive treatment from the government."

Frank Doerner, MSF's chief of mission in Guatemala, says those fears were unfounded. "It was calculated pressure, but it was not playing with the lives of the people," Doerner says of the charity's announcement that it would shut down its program. MSF earlier had successfully handed over a program in Guatemala City, Doerner notes, and MSF says it will stay longer in Coatepeque if the transition is not going smoothly. "After 5 years of being here and treating thousands of people, we showed how it was possible," says Doerner. "Now it's really up to the state to show that it's interested in taking over the responsibility that belongs to them."

–JON COHEN

HONDURAS

Why So High? A Knotty Story

Garifuna culture, discrimination against gay men, massive migration, the Cold War, and ignored prisoners all are theories that attempt to explain this country's serious epidemic

SAMBO CREEK, TEGUCIGALPA, AND LA CEIBA, HONDURAS—As a small group of men and women from this impoverished fishing village watch intently, Daniel Martínez holds up a placard that shows horrific photos of diseased female and male genitals. "Syphilis!" he yells, and the group, which is sitting under a thatched-roof shelter on the beach, looks down at what amount to bingo cards that Martínez has given them. Those who have a syphilis square mark it with an uncooked bean. The HIV/AIDS education game, *Lotería Vive*, continues with pictures of other sexually transmitted diseases and cartoons of transvestites, a drunken man, and then the Grim Reaper. "Oh!" groans the crowd at the last card, but one man has bingo and yells, "*Lotería*!" Martínez, who works with the Pan American Social Marketing Organization (PASMO), hands the winner a baseball cap and two condoms.

The residents of this village are Garifuna, so-called Black Caribs who are descendents of shipwrecked Nigerian slaves and who have maintained a distinct culture for more than 200 years. The best HIV studies done in this and three other Garifuna communities—which were conducted by the Ministry of Health more than 7 years ago—found that the adult prevalence was an astonishing 8.4%. Martínez plays *Lotería Vive* in this and other Garifuna villages in the region several times each week.

In 2005, Honduras in general had an adult prevalence of 1.5%, according to the Joint United Nations Programme on HIV/AIDS. That makes it the hardest-hit country in Central America other than relatively tiny Belize (see p. 483). The spread is mainly through heterosexual sex, which is reflected by a nearly 1:1 ratio of male to female AIDS cases. Yet the virus has also spread widely through the community of gay men, who have a prevalence of 13%—even higher than that of female sex workers, at 9.7%. By November 2005, almost 4500 people were receiving anti-HIV

Game theory. PASMO dispatches Daniel Martínez to Garifuna communities to teach HIV prevention through the bingolike *Lotería Vive*.

Published by AAAS

Mission Possible: Integrating The Church With HIV/AIDS Efforts

TEGUCIGALPA AND JUTICALPA, HONDURAS—Throughout heavily Catholic Latin America, few topics have riled those working to slow the spread of HIV more than the Vatican's opposition to condoms. Many HIV/AIDS workers have also decried what they see as the tendency by many denominations to treat as outcasts the two groups especially hard hit by the epidemic: homosexuals and sex workers. But in Honduras especially, church leaders are now trying to become part of the solution with stepped-up efforts that aim to slow HIV's spread and help the infected.

These church representatives are not, by any means, advocating the use of condoms, as Maryknoll sisters in Guatemala do with sex workers and other at-risk people they help (see p. 480). But representatives from four denominations are working with the United Nations Population Fund (UNFPA), which is famous for promoting family planning, in the year-old Interreligious Committee to contribute to Honduras's national strategic plan for confronting its HIV/AIDS epidemic. "This is the first time we've worked with faith-based organizations, and the nice thing is we put our position on the table," says Alanna Armitage, who heads the UNFPA office here. "We would not work with them if we couldn't talk about condoms or they said they weren't effective. There's no more time to fight on this."

The representatives from the Episcopal, Evangelical, Adventist, and Catholic churches do not speak with one voice about condoms; some think, for example, that they should be promoted if one partner in a marriage is HIV-infected. Nor do they exactly embrace homosexuality. "We don't have a specific program with homosexuals, but where we work, there are people with HIV/AIDS, and we treat them like anyone else," says Elvia Maria Galindo, a committee member speaking for the Episcopal church. "We're all sinners."

But Javier Medina, a gay activist here, charges that the religious community—particularly Evangelicals—have fanned the rampant homophobia in the country. He points to marches held by Evangelicals that protested the government's decision in 2004 to officially recognize his group, called Kukulcán, and two other gay organizations. "This created more hatred toward us," says Medina, adding that a few dozen gay men have recently been killed in hate crimes and that his group has received death threats. This does not reflect the opinion of other denominations, however, says Carmen Molina, the committee's Catholic representative.

Crossing the divide. Padre Alberto Gauci provides many HIV/AIDS prevention and care services in Juticalpa.

Although Padre Alberto Gauci, a Franciscan, does not condone homosexuality, he's fervently trying to help thwart HIV at a men's prison in Juticalpa, 3 hours from the capital. Gauci, who favors flip-flops, jeans, and T-shirts and looks more like an aging hippie than a clergyman, is on a somewhat quixotic quest to build a new prison in Juticalpa, where he runs an HIV/AIDS orphanage and hospice. The prison, built more than 100 years ago for 90 inmates, currently holds more than 400 men who sleep at least two to a bunk. More than 5% are known to have AIDS. In December 2005, no HIV tests or anti-HIV drugs were available. "The church has to play a role because people have lost all hope with politicians here," says Gauci, a native of Malta. "Illness is spreading in the prison in a very accelerated way."

Gauci supports his efforts by running a bakery and occasionally staging horseraces and dogfights on the grounds of his compound. "Gambling is not a sin if you're raising the money for good things," shrugs Gauci. Now that's working in mysterious ways.

–J.C.

drugs, up from 200 three years earlier. But the national AIDS committee, CONASIDA, estimates that the drugs are reaching only about one-third of those with advanced disease.

No convincing studies explain how the virus made so much headway in Honduras, but theories abound. Epidemiologist Manuel Sierra, who headed the Ministry of Health study of the Garifuna and now works at the National Autonomous University, says in most countries in the region, the virus entered through gay men and then "incubated," which means it took a long time to bridge into other communities. The first AIDS cases in Honduras were also gay men, he says, but HIV quickly spread through heterosexual sex, both in the Garifuna community and the country at large. "The main difference between Honduras and the rest of Central America is the incubation period," posits Sierra.

A key distinguishing factor in Honduras, he contends, was the country's role during the Cold War. Sierra notes that when the first AIDS cases were detected in the early 1980s, the Cold War was raging, and U.S. military personnel were flooding into Honduras in an attempt to influence the civil wars in neighboring Nicaragua, El Salvador, and Guatemala. "Honduras was the center used by the United States to fight all the countries," says Sierra. The influx of soldiers—including Nicaraguan contras who staged attacks from Honduras—led to a boom in sex workers, which in turn played a "major role," he says. César Núñez, a Honduran epidemiologist who heads the multicountry PASCA study of HIV prevalence in high-risk groups in Central America (see p. 480), says this is "a good hypothesis."

Above and beyond. Honduras has more HIV-infected patients than any country in Central America. They frequently fill the beds at Tegucigalpa's Torax Hospital.

As in other countries, prisoners are another driver of the epidemic in Honduras. A Ministry of Health study found a prevalence of 7.6% in prisons. "That's the ideal population to spread the virus," says Sierra. "You have spouse visits, lots of homosexual sex, low access to condoms, and lots of HIV." Núñez and Sierra say rampant migration has also played a central role. In particular, the country has a large num-

Published by AAAS

ber of merchant seamen, many of whom travel to Asia and Africa.

Although the Garifuna do not explain the country's high prevalence—they only number about 100,000 out of a population of 7.3 million—they are an important part of a complex story, says Sierra. When he tried to tease out why Garifuna have such a high prevalence, he found no evidence that they were more promiscuous than the *ladinos* who make up the majority in the country. Yet this has become a common belief, in part because Garifuna more openly discuss their sexual habits. "Garifuna as a group are more innocent, and they'll give you a straight answer," says Sierra. "We *ladinos* have learned how to lie."

Garifuna, some of whom make their livings as merchant seamen, also frequently migrate to the United States and other countries for work. Sierra notes that many shuttle between the large Garifuna community in New York City, which itself has a high HIV infection rate.

Garifuna have other risk factors, including widespread poverty and less access to health services. The culture also has many myths that make it more difficult for HIV-prevention educators. "They believe a spirit can enter a person and therefore that HIV is an inherited thing," says PASMO's Martínez, who is half Garifuna himself. "And when a person is showing symptoms, they think it's an ancestor asking for a religious ceremony."

Sergio Flores, the top HIV/AIDS doctor in La Ceiba—the nearest city to Sambo Creek—worries about highlighting the high prevalence in the Garifuna, because the population already suffers so much stigma and discrimination. "The community was essentially forgotten about, but when HIV arrived, we put our eyes on them," says Flores. "It doesn't seem right to me. And if you go to the street and ask the people about AIDS issues, many of them think 'AIDS, it's not in my house—it's the house of the Garifuna.' " **–JON COHEN**

BELIZE

Taking It to the Streets

An unusual prevention program targets gang members, who are seen as particularly vulnerable to HIV

BELIZE CITY, BELIZE—Shortly after Douglas Hyde started working 4 years ago doing HIV/AIDS prevention work with gang members, he was welcomed with a "pint bottle" to his face that left a nasty scar above one eye. Today, Hyde, a former gang member, continues the work through a multipronged government program called Youth for the Future that attempts to link violence reduction with HIV/AIDS education.

As Hyde drives around the rough South Side streets where he grew up, he repeatedly toots the horn of his van at gang members. "What's up, fam?" he asks a group of men and boys hanging out on one street who don't exactly look like his family. The group gives a warm "Ya ya" to "Dougie," who has o-n-e l-o-v-e inked across his fingers and barbed wire tattooed on a bicep. Several of the men wonder whether he has leads on any jobs. "I have become the job god in the street," says Hyde.

This is Blood territory, the gang that Hyde used to run with until a showdown with the rival Crips scared him straight, and he notices the finer details of the street. The pile of used clothing for sale on the sidewalk is a front for dealing drugs. Most of the guys in this group are "strapped" with pistols. "Scopes" at second-story windows of the incongruously colorful clapboard homes are monitoring his every move. And he sees something else that may be less than obvious to outsiders: a strong link between the gang lifestyle and Belize's high prevalence of HIV, which at the end of 2005 had infected 2.5% of adults. That's why Youth for the Future believes that finding people legitimate jobs and encouraging them to quit gangs is a potentially powerful HIV prevention strategy.

Although many Latin American countries have problems with gangs, a 2005 report by the nonpartisan U.S. Congressional Research Service said "the largest and most violent" ones are in Central America and Mexico. According to the report, several factors have led to an increase in gangs: weapons left over from the many civil wars in the region, the stepped-up U.S. deportation of law-breaking immigrants, and staggering income inequalities in Belize and its neighbors. Youth for the Future is one of the few efforts that explicitly targets gang members as "at-risk youths" for HIV infection.

Ganging up on HIV. Youth for the Future's Douglas Hyde (*right*) found these former gang members jobs with a company that's clearing this junkyard.

Not only do gang members often share one woman, Hyde says, but "transactional sex" for a meal or protection is also the norm. "Give some, get some," says Hyde. Condom use is also low. "And some guys in the street, especially the leaders, believe that they don't need to take the HIV test," says Hyde. "They believe they just need to send their girls or wives to take the test to know their status. We're telling them that's not true."

Supported by the United Nations Population Fund and a grant from the OPEC Fund, Youth for the Future maintains a resource center that's essentially a hangout for anyone, and gang members are welcome. It stages frequent HIV/AIDS prevention education sessions and has a big bowl filled with free male and female condoms, free pamphlets on HIV/AIDS prevention, and Internet access for a small fee (free to students). "They have done tremendous work," says epidemiologist Paul Edwards, head of the Ministry of Health's National AIDS Program. "These kids have a lack of education and don't make the best decisions possible."

No study has ever assessed HIV prevalence in gang members in Belize, which has a tiny population of 280,000 people. A study done in the country's one prison—which almost every longtime gang member knows intimately—found an HIV prevalence of 4.6%. Youth for the Future plans to start offering HIV counseling and testing, and Hyde hopes to recruit gang members to participate in a prevalence study. Meanwhile, he's become increasingly cautious about how he conducts his business. "I'm good with everyone," says Hyde. "But I'm very smart now to recognize when I shouldn't be around."

–JON COHEN

Published by AAAS

South America

With its bold 1996 policy to offer top-of-the-line AIDS drugs to everyone in need, Brazil catalyzed the "universal access" movement. Spurred by AIDS activists and donors, many governments in South America have followed suit. Although prevention has stumbled in many countries, Brazil, Peru, and Argentina each have had innovative campaigns, and they have also supported cutting-edge HIV/AIDS research. In part because of these efforts, the epidemic has not spread far beyond high-risk groups, although there's increasing evidence of "bridging" to the general population.

	Est. HIV/AIDS cases	Est. adult prevalence (%)	Population
Argentina	130,000	0.6	39,921,833
Bolivia	7,000	0.1	8,989,046
Brazil	620,000	0.5	188,078,227
Chile	28,000	0.3	16,134,219
Colombia	160,000	0.6	43,593,035
Ecuador	23,000	0.3	13,547,510
Guyana	12,000	2.4	767,524
Paraguay	13,000	0.4	6,506,464
Peru	93,000	0.6	28,302,603
Suriname	5,200	1.9	439,117
Uruguay	9,600	0.5	3,431,932
Venezuela	110,000	0.7	25,730,435

TABLE SOURCE: UNAIDS/CIA WORLD FACT BOOK

BRAZIL

Ten Years After

After stunning the world by offering antiretroviral drugs to all in need, this country is struggling with the escalating costs of providing free HIV/AIDS care

RIO DE JANEIRO AND SÃO PAOLO, BRAZIL—In 1996, when it first became clear that potent cocktails of anti-HIV drugs could dramatically extend the life of an infected person, the $15,000-a-year price tag seemed out of reach to all but the world's wealthiest people. Brazil, which already had a progressive prevention program, said to hell with that. A middle-income country with more HIV-infected people than any other in Latin America or the Caribbean, Brazil declared that it would provide the treatment, at no charge, to every resident who needed it. And the government would bankroll this seemingly outlandish promise in part by having Brazil's own drugmakers produce copies of antiretroviral drugs that major pharmaceutical companies had patented.

Brazil soon became a poster child for the access movement, which argues that everyone, everywhere can have antiretroviral drugs by purchasing knockoffs—outside Brazil, mostly made by generic drug companies in Asia—and by hard bargaining with Big Pharmas. By the end of 2005, 1.3 million HIV-infected people in poor and middle-income countries were receiving steeply discounted drugs, up from 240,000 in 2001. Brazil today has 180,000 people on antiretroviral drugs; 20% are made in the country, and the rest are purchased from Big Pharmas—typically after the government stages heated, much publicized, negotiations to exact price breaks.

As aggressive as Brazil has been about confronting Big Pharma, a growing number of insiders are criticizing the country for going soft and too readily acceding to Big Pharma's wishes. Brazil manufactures only eight antiretroviral drugs, all of them older preparations. Fourteen newer drugs offer many advantages, such as fewer side effects, more potency, and effectiveness against many drug-resistant viruses. Although Brazil has repeatedly threatened to break patents and make copies of these newer drugs, each time push has come to shove, government officials have backed down and cut deals with the Big Pharmas that have made some leading Brazilian AIDS researchers and activists blanch. "This has been a huge disappointment for us," says Pedro Chequer, who twice headed Brazil's national AIDS program and now works for the Joint United Nations Programme on HIV/AIDS (UNAIDS). Alexandre Granjeiro, another former head of the AIDS program, says Brazil must violate patents and risk incurring the wrath of Big Pharma and other industries that hold fast to intellectual-property regulations. "It's important to the world," says Granjeiro, who is director of the São Paolo State Health Institute. "If we make this ball roll here, it will make the ball roll everywhere."

Turnaround?

In 1992, the World Bank predicted that Brazil would have 1.2 million infected people by 2000. But because Brazil meshed aggressive prevention efforts with its pioneering treatment program, this dire prediction has not come true. According to UNAIDS estimates, at the end of 2005, 620,000 Brazilians were infected with HIV. The adult HIV prevalence in the country is a modest 0.5%, but because it is the most populous country in Latin America with 188 million residents, Brazil still accounts for more than one-third of the HIV/AIDS cases in the region.

As in North America and Europe, AIDS first surfaced in Brazil in upper-middle-class gay men, many of whom were politically active in the democracy movements that blossomed when 2 decades of military rule

Published by AAAS

ended in 1985. "The community movement became extremely well organized, more than in the United States," says Ezio Tavora dos Santos Filho, a prominent AIDS activist who learned of his infection that year. In 1988, when Brazil rewrote its constitution, it declared that health care was a right, and 3 years later, the country offered HIV-infected people free AZT—then the only antiretroviral drug on the market.

By 1992, the virus had spread far and wide, with equal numbers of AIDS cases that year occurring in gay and bisexual men, heterosexuals, and people who injected cocaine—but still, it did not take off to the degree once feared. It's difficult to untangle precisely why, although Chris Beyrer, an AIDS epidemiologist at Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland, and co-author of a 2005 World Bank case study of Brazil, credits aggressive prevention campaigns. The Ministry of Health alone tripled the number of condoms it distributed between 2000 and 2003, the report notes, and government and nongovernmental organizations alike boldly reached out to gay men, sex workers, and injecting drug users.

Other factors contributed as well, says Beyrer. Antiretroviral treatment lowers the level of virus, likely making recipients less infectious. And the availability of treatment encouraged people to undergo HIV tests, which in turn can lead those who are infected to take more precautions. A change in drug-use trends—injecting cocaine largely fell out of fashion as many users switched to smoking the drug—contributed to the declining spread of HIV, too. "Brazilians hold on to how severe their epidemic is, but the bottom line is it could have been much worse," says Beyrer. And because Brazil controlled HIV's spread early on, he says, it made offering state-of-the-art treatment to everyone in need much more feasible.

Provocative prevention. Gabriela Leite heads Davida, an NGO for sex workers that launched a clothing line to raise money to fight the spread of HIV.

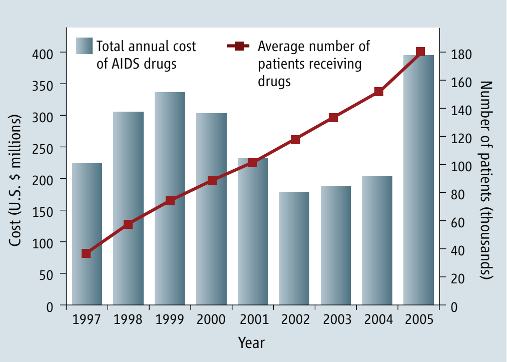
SOURCE (INSET): MINISTRY OF HEALTH, BRAZIL

State pharma. Farmanguinhos head Eduardo Costa hopes to ramp up production of antiretroviral drugs at the company's new high-tech plant.

Rights and wrongs

Brazil became an icon for HIV-infected poor people everywhere—and a punching bag for critics—following its 1996 decision to offer its residents cocktails of three antiretroviral drugs that had just become available. One of the strongest naysayers was the World Bank, which by then had committed a whopping $750 million to help Brazil combat its AIDS epidemic. "We received a lot of pressure to not implement combination therapy," remembers Valdiléa Veloso, who now directs the Evandro Chagas Clinical Research Institute at Fundação Oswaldo Cruz (Fiocruz), a biomedical research center run by the Ministry of Health. Formerly with the national AIDS program, Veloso says bank representatives urged them to put more money into prevention instead. "They all argued it was a crazy decision to offer triple therapy in Brazil because of the complexity, the cost," she says.

Objections came from within the country, too. "I was very skeptical," acknowledges Mauro Schechter, a leading AIDS researcher at Federal University in Rio de Janeiro. Because of limitations in the country's health care infrastructure and clinician training, Schechter worried that many infected people would not adhere to the complicated treatment regimens, leading to widespread drug resistance. "I was obviously wrong," says Schechter now. Brazil's Ministry of Health reports



that between 1996 and 2002, AIDS mortality dropped 50%, and an estimated 90,000 deaths were averted. The government says it saved $1.2 billion that would have been spent on hospital admissions and treating the opportunistic infections of AIDS.

Nor have the disaster scenarios of the rapid spread of drug-resistant strains come to pass. "We don't have any evidence of primary resistance increasing," says Amilcar Tanuri, who runs a molecular biology lab at Fundão Isla in Rio, a branch of the Federal University, referring to the spread of resistant strains between individuals. Yet Tanuri notes that "secondary" drug resistance, which develops while on treatment, is becoming more widespread, requiring many to change their medicines. "There's no way around it," he says. Combine that with the growing number of people on treatment, and Brazil is now faced with importing an increasing quantity of ever-more-expensive drugs. "The cost of treatment is going up and up and up," says Tanuri. More people on treatment also means more work for already-overstretched clinics. "Brazil has not done the homework over the past 10 years," complains Schechter, who would like to have seen the government use research to assess how best to use its limited resources. "I'm really concerned about the sustainability of the program."

Published by AAAS

Free Drugs ≠ Quality Care

RIO DE JANEIRO, BRAZIL—Thanks to the persistence of a niece, Luis Silva, 50, made his way to the highly regarded AIDS clinic at the Evandro Chagas Clinical Research Institute one morning in June. After suffering persistent fevers and night sweats, Silva in August 2005 had sought medical care at a clinic near the poor neighborhood where he lives. An HIV test indicated that he had been infected, but Brazilian regulations require a second, confirmatory test before doctors order expensive immune tests, which in turn are needed before they can prescribe antiretroviral drugs. The doctors treated what they thought was a pulmonary infection, and for a time Silva's condition improved, so he skipped the second test. But then the slightly built man lost 20 kilos and developed a hacking cough, which led him to several other doctors, who offered little help. Finally, his niece, who is a nurse, brought him here.

A chest x-ray taken that day showed strong evidence of tuberculosis, and Silva's doctor said she was all but certain that he has AIDS. Still, even she had to wait 10 days for the lab to determine his HIV status, as only pregnant women have access to the rapid test that can give results in a few hours. The clinic's director, Valdiléa Veloso, notes that many other facilities in Brazil routinely run out of HIV test kits. "It's crazy," says Veloso. "It would have been much better for the government to have made the decision about rapid tests years ago."

Delayed reactions. Luis Silva (*left*) had to jump through many hoops to see whether he was HIV infected and eligible for treatment.

As progressive a stance as Brazil has taken on HIV/AIDS prevention and care, it remains a middle-income country offering uneven health care services. "In Rio, it's not uncommon to receive in the emergency room HIV-infected people who were not treated," says Pedro Chequer, who twice headed the country's national AIDS program and now works for the Joint United Nations Programme on HIV/AIDS. "The health care system here is collapsing."

Activist Ezio Tavora dos Santos Filho recently completed a report of the tuberculosis care offered in Brazil, which he notes is in the "shameful position" of being 15th on the World Health Organization's list of 22 countries that have a high TB burden. "It's indefensible," says Tavora. According to his report, federal, state, and city TB programs are only now beginning to work together, as officials recognize that 12% of HIV-infected people are coinfected with TB.

Solange Cesar Cavalcante, who heads the TB program for Rio, notes that unlike HIV/AIDS, TB is not a "sexy" topic and so far has not mobilized affected communities. Says Cavalcante, "Tuberculosis is trying to learn from the AIDS program." –J.C.

Tripping on TRIPS

Between 1997 and 2004, the average annual cost of antiretroviral therapy in Brazil dropped from $6240 per patient to $1336. That decline allowed the country to treat more people without increasing its budget for AIDS drugs. But because Brazil has steadily purchased more imported drugs, in 2005 the per-patient annual cost jumped to $2500 (see graph, p. 485). Forecasts suggest that costs will continue to climb unless the country violates patents or negotiates better deals with Big Pharma.

At the crux of Brazil's current dilemma are the World Trade Organization's patent rules, known as the Trade-Related Aspects of Intellectual Property Rights (TRIPS). In 1996, when Brazil decided to offer HIV cocktails, it passed a law that enforced the TRIPS agreement. The new regulation meant that Brazil could legally produce anti-HIV drugs patented before the signing—but not the improved antiretroviral drugs and new classes of drugs that have come to market over the past 10 years. Today, Brazil's Ministry of Health spends 80% of its $445 million annual budget on imported antiretroviral drugs. And the ministry estimates that between 2006 and 2011, the annual cost of purchasing just three of these drugs—Merck's efavirenz, Abbott's lopinavir/ritonavir, and Gilead's tenofovir—will jump from $145 million to $248 million.

If the government instead made the drugs at the state-owned pharmaceutical company Farmanguinhos, the ministry says the country would save $769 million over that period. "If there's no change in the price of second-line drugs, no country like Brazil will be able to afford them," says Luiz Loures, a Brazilian epidemiologist who works at UNAIDS.

"Brazil has the technical capacity to produce all of the drugs," says Paolo Teixeira, who ran Brazil's AIDS program from 2000 to 2003 and now works as a consultant for São Paulo's AIDS program. And he says that gives the country a strong negotiating tool when purchasing antiretroviral drugs in bulk from Big Pharmas. Essentially, the government has said, "If we don't like your price, we'll violate the patent and make the drug ourselves." This is allowed under the TRIPS agreement, which says signatories can invoke what is known as a "compulsory license" to address public health emergencies. No country has yet done so, however, because of fear of damaging international trade relations. Brazilian President Luis Inácio Lula da Silva twice has promised to use the compulsory-license clause for anti-HIV drugs but has backpedaled both times, complains former AIDS program head Chequer. "They were cowards by not doing that," says activist Tavora. "That could be very useful to all of us, to the whole world."

David Greeley, Merck & Co.'s spokesperson for Latin America, says if Brazil invokes compulsory licensing, it will ultimately harm the people the government is trying to help. "We've tried to convey to our counterparts in Brazil that it's not in the long-term interest for Brazil to adopt this stance," says Greeley. As with other Big Pharmas, Merck invests in research and development of new products because intellectual-property regulations exist, he says. "Intellectual property is an incentive to innovation, not a barrier to access," he maintains.

Retaining the lead

In the Rio suburb of Jacarepaguá, there are clear signs that the government once again wants Brazil to lead the charge against Big Pharma with more than rhetoric. Jacarepaguá's Estrada dos Bandeirantes has long housed the gleaming offices of international giants such as Abbott and Roche, both of which have crossed swords with Brazil over pricing of their anti-HIV drugs. In August 2005, a new resident moved into the neighborhood: Farmanguinhos, the government-owned drugmaker.

Farmanguinhos's new factory, once owned by GlaxoSmithKline, has five times the pro-

Published by AAAS

Patently absurd. Not invoking compulsory licenses is deadly, says Pedro Chequer.

duction capacity of its old plant on the other side of the city. Company Director Eduardo de Azeredo Costa has ambitions beyond just manufacturing more antiretroviral drugs. He says Brazil needs to start producing the active pharmaceutical ingredients used to make the drugs, which it now purchases from India and China. Costa says these are often of inferior quality, so by making its own, Farmanguinhos can both reduce costs and avoid expensive delays in production.

But even with these changes, making the new generation of antiretroviral drugs will be challenging for Brazil. "It's a lie that if we had no patents, we just can from right today produce generic medicines for all drugs," says epidemiologist Francisco Basto, a leading AIDS researcher at Fiocruz. "This will be a very, very complicated issue for the coming few years."

Costa agrees but says Farmanguinhos and other drugmakers must rise to the occasion, for the sake of Brazil and other cash-strapped countries. As Costa walks around the plant's new high-tech machines—several of which are still wrapped in plastic—he notes that representatives from two dozen countries have toured the facility in hope of following in the Brazilian government's footsteps. "People of the world want us to be much better than we are," says Costa. "We have to answer to this demand."

–JON COHEN

ARGENTINA

Up in Smoke: Epidemic Changes Course

Over the past few years, HIV infections of heterosexuals have eclipsed those of injecting drug users and gay men

BUENOS AIRES, ARGENTINA—Stella Maris Todaro is part of a battalion of *promotorios* hired by the government to educate their communities about HIV/AIDS. "I started this work 15 years ago because I saw my children were addicted, shooting drugs," says Maris, who lives in a poor neighborhood called a *villa miseria*. Whereas most countries in Latin America then had AIDS epidemics concentrated in homosexual men, Argentina, like its neighbors in the Southern Cone of South America, had an equally large problem in injecting drug users (IDUs) who shot cocaine. As it turned out, Maris's two sons both became infected by sharing syringes and died from AIDS. Although she was not an IDU herself, a sometime partner was, and in 1995, Maris learned that she, too, was HIV-positive.

Today, Maris, 52 and a grandmother, better characterizes the average HIV-infected person in Argentina than do her sons. In a dramatic shift seen across the Southern Cone, IDUs largely have either died from AIDS or stopped injecting cocaine and switched to smoking the much cheaper *pasta base de cocaine*, or *paco*, a low-grade paste. "We have a great change of the use of drugs in Argentina," says epidemiologist Claudio Bloch, head of the HIV/AIDS program for the city of Buenos Aires. Bloch, like many other experts, contends that *paco*'s rise in popularity is a result of "the crisis," the sharp devaluation of the peso that occurred in 2001 and 2002, although the same shift has occurred in other Southern Cone countries that did not suffer an economic collapse.

By December 2005, HIV had infected 130,000 people in Argentina, or 0.6% of all adults, a percentage that has remained steady for several years. Ministry of Health figures from 2004 show that 50.7% of the people with AIDS had been infected through heterosexual sex, whereas men who have sex with men (MSM) accounted for only 18%, and IDUs were at 16.6%. A similar analysis from 1982 to 2001 shows that 40.1% of the AIDS cases were IDUs—more than either MSM or heterosexuals. In Buenos Aires, the evidence is more telling still: IDUs accounted for only 5.2% of the new infections between 2003 and 2005. Now, says Bloch, the new infection rate in men and women is almost the same. "The heterosexualization of the epidemic is so strong," he says.

As more women become infected, Maris's services become increasingly valued. "I've learned a lot of things from Stella," says Sara Tapia, 33, a mother of four who also works as a *promotorio*, lives in a *villa miseria*, and is HIV-positive. "In life, we have to be what we are. We mustn't pretend. We're always going to be that." One of Tapia's most difficult challenges, she says, is that her husband refuses to get tested: "It's not something he

Cold truth. HIV/AIDS in Argentina is increasingly a disease of poor women such as Sara Tapia (*left*), a mother of four who lives in this *villa miseria*.

Published by AAAS

wants to talk about, and it's obviously painful for him, so we don't dwell on it."

Drug drop. Claudio Bloch's group has documented a steep decline in HIV spread via shared needles.

Great expectations

Argentina was one of the first countries in Latin America to offer antiretroviral drugs to everyone in need, but it has not received the worldwide praise that's been poured onto neighboring Brazil for making a similar commitment. "People talk about Brazil because the Brazilians have done a very good job of marketing what a very good job they've done," says Pedro Cahn, a leading AIDS researcher in Buenos Aires who heads the Fundación Huesped and is chief of infectious diseases at Hospital Juan Fernández. But he also stresses that Brazil has a "more consistent" national program in many ways.

Both of Maris's boys became sick before potent cocktails of anti-HIV drugs had come to market, but she was luckier. Today, the virus is not detectable in her blood, and her immune system is robust. Tapia similarly is doing well on a cocktail of drugs.

Some 30,000 infected people in Argentina are currently receiving treatment, which the government says is 100% of those with advanced disease. Mother-to-child transmission, which anti-HIV drugs can prevent, has dropped to 3%. "It's similar to Paris," notes Bloch.

Yet many AIDS researchers and patients complain that the government program has many shortcomings compared to wealthy countries. That is a central dilemma for Argentina, which long has seen itself as the most European country in Latin America, yet frequently—especially since the crisis—finds itself with rich-country expectations but poor-country limitations.

One of the biggest problems is that government clinics and hospitals are short staffed. "You have to wait for everything," says Roxana González Montaner, a clinician who works in a poor part of the city. She notes that there are long lines every morning, and that many doctors here work in both public and private practice to make ends meet. Lab tests require more long waits, and the results often do not arrive back at clinics for weeks or even months. "We can make many things happen for [some people] but not for everyone," says González.

Pharmacies all too frequently run out of anti-HIV drugs. "This morning, we didn't have abacavir at my hospital," says Cahn, referring to an increasingly popular drug for people starting treatment. "Ask me why, we don't know."

Carlos Zala, an AIDS clinician and researcher at Hospital Juan Fernández, says the government needs to spend more money on monitoring treatment. "HIV [care] is much more than just providing antiretroviral drugs," says Zala, noting that it's often difficult for people to learn their immune status or the levels of HIV in their blood. He also faults the government for not monitoring the treatment program itself, which his team is now starting to do by carefully following a cohort of treated people to gauge the emergence of drug resistance and health problems. "This is typically Argentina: a good thing, a good action, that no one is controlling," says Zala. "We will provide medication, but no one will see whether it works."

–JON COHEN

PERU

A New Nexus for HIV/AIDS Research

Talented investigators and explosive spread in men who have sex with men have made this country a hot spot for clinical studies

LIMA, IQUITOS, AND NAUTA, PERU—On a Friday night this June at a gay disco in Iquitos, a jungle city that's the jump-off point for touring the Amazon rainforest, drag queens danced to the thump of "*Voulez-vous coucher avec moi?*" in a Miss Adonis contest. The event, staged by the Asociación Civil Selva Amazónica, was part entertainment, part HIV prevention, and part recruitment for an AIDS vaccine trial.

Welcome to Peru, a somewhat incongruous hotbed of HIV/AIDS research. "Everyone's going to Peru, and it's not because they have a huge epidemic," says Robert Grant, a virologist at the University of California, San Francisco (UCSF), who runs one of many collaborative projects now under way. "It's because of the research climate."

Intensive efforts are now under way to understand the country's perplexing epidemi-

Recruiting station. "Lashmi" leads a teach-in about drag queens that doubles as an attempt to find volunteers for an AIDS vaccine trial.

Published by AAAS

ology—the epidemic is concentrated among men who have sex with men (MSM) and has not "bridged" much to other groups—and to evaluate new treatment and prevention strategies. The scope and scale of the research enterprise is especially remarkable given the government's foot-dragging when it comes to offering anti-HIV drugs to people who need them (see sidebar, right).

Only 0.6% of Peruvian adults were infected with HIV by the end of 2005, according to the Joint United Nations Programme on HIV/AIDS (UNAIDS). But studies suggest that the prevalence in Peruvian MSM—a group that includes many bisexuals who consider themselves heterosexual—is 10% in Iquitos and the surrounding area and more than twice as high in Lima. It's on this group that researchers have focused their attention. "It's a very concentrated epidemic, and we have a very good relationship with the community," explains epidemiologist Jorge Sánchez, who runs Asociación Civil Impacta Salud y Educación (Impacta), a nongovernmental organization based in Lima.

Similarly, Carlos Cáceres, an epidemiologist at the Universidad Peruana Cayetano Heredia in Lima, has a team of AIDS researchers working closely with high-risk communities to evaluate behavioral interventions, viral spread, and strategies to reduce stigma and discrimination. "There's a lot to be studied here," says Cáceres.

Both Sánchez's and Cáceres's groups have strong ties to U.S. academics, participate in international multisite studies, and receive substantial funding from the U.S. National Institutes of Health (NIH). A challenge, says Cáceres, is ensuring that such collaborations serve both Peru's own interests and those of the funder.

Why Peru?

Many factors have contributed to Peru becoming a nexus of collaborative HIV/AIDS research, but explanations usually return to Sánchez and Cáceres. "There are great people here," says Rubén Mayorga, the Lima-based UNAIDS country coordinator. "And there's an acknowledgment that HIV is a big problem among gay men or men who have sex with men."

Sánchez and Cáceres—who, to the frustration of many, have a strained relationship—command wide respect from colleagues around the world. Sánchez was the first of some 40 Peruvian researchers who were funded by NIH's Fogarty International Center to train at the University of Washington (UW), Seattle, with King Holmes, a renowned expert on sexually transmitted diseases. Sánchez then headed Peru's national AIDS program within the Ministry of Health. When he left, he took many members of his team and started Impacta. His group now collaborates with both UW and Grant's lab at UCSF. Cáceres has a doctorate in public health from UC Berkeley and works closely with Thomas Coates's AIDS research team at UC Los Angeles.

Mayorga says Sánchez and Cáceres have a deep understanding of the communities that they are studying because they are both part of them. "I know exactly what it means to have a partner who weighs 40 kilos and you need to take him to shower because he cannot shower himself," says Sánchez, who had a partner die of AIDS in 1990. "I cannot take my personal life out of my thinking." Cáceres, too, says his personal links to the community shape the way he does epidemiology. "It's public health and prevention mixed with sexual rights and human rights and empowering the community," he says.

Epidemiologist Javier Lama, a co-investigator with the NIH-sponsored HIV Vaccine Trials Network, says Peru is particularly poised to do prevention studies because of the high incidence, or rate of new infections, in MSM. Such high incidence rates, ranging from 3.5% in Iquitos to 6.2% in Lima, enable researchers to discern whether a prevention intervention works with relatively smaller, shorter trials

Universal Access: More Goal Than Reality

Late stage. Milton Ramírez needs antiretroviral drugs, but he must wait for test results before he's eligible.

LIMA AND IQUITOS, PERU—As much as Peru has taken a leading role in conducting HIV/AIDS research, the government has lagged when it comes to offering antiretroviral treatment to infected people. Peru didn't begin providing free antiretroviral treatment to all in need until 2004—8 years after neighboring Brazil—and did so only after being prodded by a grant from the Global Fund to Treat AIDS, Tuberculosis, and Malaria. "They have pushed us to work faster," acknowledges Pilar Mazzetti, the minister of health. "We've taken a long time to have a response."

Some 7000 people now receive anti-HIV drugs in Peru, up from 2000 a mere 2 years ago. Robinson Cabello, who runs the Via Libre clinic in Lima and in years past helped his patients sue the government for access to anti-HIV drugs, says up to 20% of people who need antiretroviral drugs immediately still do not receive them. And outside Lima, which is home to about 70% of the infected people in the country, the problem is especially acute.

Take Iquitos, a jungle city in the north of the country that has a high HIV prevalence in men who have sex with men. The main hospital has repeatedly run out of anti-HIV drugs for the 110 people receiving the treatment. "The last 2 months, we didn't have enough drugs to support our patients," says Cesar Ramal Sayag, head of infectious diseases at the Regional Hospital of Loreto. Sayag says he also has to wait several weeks to receive results of tests for CD4 white blood cells—which must be air-shipped to Lima—and that government rules do not allow him to start patients on treatment without that information. "The national program will continue this way for 10 years, and they won't change," says a frustrated Sayag.

Across town at the Hogar Algo Béllo, a hospice run by a Catholic priest, a 22-year-old gay man named Milton Ramírez is suffering from untreated late-stage AIDS. Ramírez has been ill for 2 years. And although two separate tests have confirmed his HIV infection, his blood was drawn to measure his CD4 cells just a few weeks ago, and his doctors are still waiting for results before they can treat him.

Marco Calixtro, a doctor in town at Asociación Civil Selva Amazónica, is part of the team that cares for Ramírez and other patients at the hospice. "It's pathetic," Calixtro says. Calixtro of course knows all about the government's promise to provide antiretroviral drugs to everyone in need. But, he says, "when we look at a problem like Milton, it seems like all this stuff we hear isn't actually real."

–J.C.

than would be needed in locales with, say, 1% incidence.

Grant is now working with Lama, Sánchez, and other Impacta researchers to launch one of the most ambitious—and contentious—prevention studies in the world: an evaluation of whether antiretroviral drugs used to treat infection can lower transmission rates if *uninfected* people take them each day. Four studies of so-called pre-exposure prophylaxis (PrEP) have been blocked or aborted in Africa and Asia because of community protests about trial designs as well as problems with data quality. But Grant is confident that the placebo-controlled trial—which is slated to start in November and will test a combination of the anti-HIV drugs tenofovir and FTC in 1400 Peruvian and Ecuadorian MSM—will fly. "The advantage of working here is they have a mobilized population," says Grant. He says Peru also has a proven track record of quickly enrolling volunteers.

Leading lights. Carlos Cáceres (*left*) and Jorge Sánchez run two separate HIV/AIDS research programs in collaboration with U.S. research teams.

In addition to the PrEP study and trials of experimental AIDS vaccines, Impacta is also playing a leading role in two multicountry studies that are evaluating whether the drug acyclovir can help people infected with herpes simplex virus 2 avoid acquiring or transmitting HIV. Impacta is part of an NIH network that tests new HIV treatments, too.

Cáceres and his co-workers spend about half their effort on a multicountry behavioral study funded by the U.S. National Institute of Mental Health that's testing "diffusion of innovation" theory. The researchers identify popular opinion leaders in various poor neighborhoods, educate them about HIV prevention, and then assess whether that intervention helps lower HIV incidence in the community. This team also has a study under way to gauge whether art can reduce stigma and discrimination against HIV-infected people. On World AIDS Day last year, they distributed T-shirts made by artists to all the staff and patients at three Lima hospitals. The T-shirts had messages on them that, roughly translated, said all of us are living with HIV.

Why mainly MSM?

Although all Peruvians may be living with the HIV epidemic, the virus has not made many inroads outside the MSM population. Female sex workers, for example, have a prevalence of less than 2% in Lima, and a 2002 study of nearly 4500 sex workers from 24 smaller cities found a prevalence of only 0.62%. The prevalence in women in general is a mere 0.2%

These findings might suggest that few MSM have sex with women, but that's not the case. "A big part of the MSM community is married," says UNAIDS's Mayorga. Indeed, a survey, now in press, of more than 4000 MSM between 1996 and 2002 in Peru found that in one year, 47% of the men reported having had sex with a woman.

Cáceres suggests that the heterosexual epidemic has not taken off in part because monogamy is the norm in the Peruvian women who become infected by bisexual partners. Says Cáceres, "The epidemic stops in them and doesn't spread." He notes, too, that Peru has no injecting drug use, which in other countries is another way that the epidemic commonly bridges into heterosexual women. Sánchez says "of course it surprises me" that more women are not infected, but his work suggests that bisexual men, because of their sexual practices (typically "insertive" rather than "receptive" in anal sex), have a lower HIV prevalence than that of men who exclusively have male partners.

Net gains

A team from Selva Amazónica recently drove a few hours to the town of Nauta to attend a volleyball game. In Peru, volleyball has long had the reputation of being a sport for gay men—macho men play soccer—and the Selva Amazónica team wanted to see whether they might recruit volunteers for one of Impacta's prevention trials.

Although gay men once feared playing volleyball in public, onlookers filled the town square in Nauta to watch two teams spike the net in the sweltering Amazonian sun. "The environment for gay people in Peru has markedly changed in the last 5 years, and it's really because of the way the AIDS epidemic has been addressed," said Grant, who had come along for the ride. So far, in Nauta, however, AIDS has not had much impact: The head of the town's gay organization says he does not know anyone here who has died from the disease or is even infected.

Then again, Nauta has all the ingredients needed for HIV to take off. The only place to buy condoms this day is the town's hospital, which gives them away for family planning but charges everyone else. No one offers HIV tests. And judging by the turnout at the volleyball game, there's a substantial MSM population.

All of which explains why Selva Amazónica came here—and why Peru is so enthusiastic about research. Anyone who joins the group's studies receives free condoms, HIV tests, counseling, checkups, and education. And that means that the abundance of HIV/AIDS research here may have a huge payoff, regardless of whether the trials ultimately yield positive results.

–JON COHEN

Net gain. Volleyball games like this one in Nauta are popular hangouts for gay men, making them key sites for HIV/AIDS researchers who do prevention work and stage clinical trials.

30,000 Years of Cosmic Dust in Antarctic Ice

Gisela Winckler[1]* and Hubertus Fischer[2]

About 40,000 tons of extraterrestrial matter fall to Earth each year. A fraction of the cosmic dust is archived in the polar ice sheets [e.g., (*1*, *2*)], side by side with the much more common terrestrial dust. In addition to its astrophysical importance, cosmic dust has the potential to constrain rates of sedimentation in various geological archives. However, the cosmic dust flux, especially its short- and long-term temporal variability, is poorly known. Here, we present a high-resolution, glacial-to-interglacial record of cosmic dust flux from an Antarctic ice core.

We used ^{3}He, the rare isotope of helium, to trace the fraction of cosmic dust that retains its gas load during atmospheric entry. The samples, from the EPICA (European Project for Ice Coring in Antarctica) ice core drilled in Dronning Maud Land, cover the time period from 6800 to 29,000 years before the present. We developed a technique to sample the excess water stream of a continuous chemical meltwater analysis, allowing us to process sufficiently large ice samples (~5 kg). Particulate dust was collected on silver filters (*2*); helium isotopes were determined by mass spectrometry (*3*). Each sample covers between 300 and 600 years for the glacial and between 150 and 200 years for the interglacial.

The filtered particles contain a binary mixture of extraterrestrial and terrigeneous helium, bound in the cosmic and terrestrial dust, respectively. The helium isotopic ratios of these two end-members differ by four orders of magnitude (^{4}He/^{3}He$_{ET}$ ~ 4200 and ^{4}He/^{3}He$_{TERR}$ ~ 2.5 × 10^7), so identification of the extraterrestrial component is unambiguous.

The low ^{4}He/^{3}He ratios (Fig. 1A) measured in the bulk particulate matter are very close to that of the extraterrestrial end-member, indicating that nearly all the ^{3}He in the ice is of extraterrestrial origin. By using the reconstructed snow accumulation rates for the ice core, we derived the extraterrestrial ^{3}He flux (Fig. 1B). The ^{3}He flux is well defined at 7.5 × 10^{-13} ± 2.9 × 10^{-13} cm^3 STP cm^{-2} ky^{-1} (median ± median of the absolute deviation from the median, where cm^3 STP is cubic centimeter at standard temperature and pressure and ky is 1000 years) despite the scatter in the ^{3}He fluxes, which is caused by the small number of interplanetary dust particles (IDPs) in each sample. Most importantly, we do not observe any significant change in the ^{3}He flux from glacial (>13 ky: 7.5 × 10^{-13} ± 2.6 × 10^{-13} cm^3 STP cm^{-2} ky^{-1}) to Holocene (<13 ky: 7.7 × 10^{-13} ± 3.3 × 10^{-13} cm^3 STP cm^{-2} ky^{-1}) conditions. This relatively constant ^{3}He flux rules out the input of interplanetary dust as a driver of the late Pleistocene 100-ky glacial cycles (*4*), as previously suggested (*5*).

Our high-resolution record is in agreement with previous estimates of the extraterrestrial ^{3}He flux derived from low-latitude marine sediment cores over the past 200 ky (*6*) and from Holocene ice from Vostok (*2*), thus indicating a globally uniform deposition of ^{3}He-bearing IDPs. This supports the use of ^{3}He as constant flux proxy in paleoclimate studies, for example, to derive quantitative accumulation rate estimates in deep ice cores.

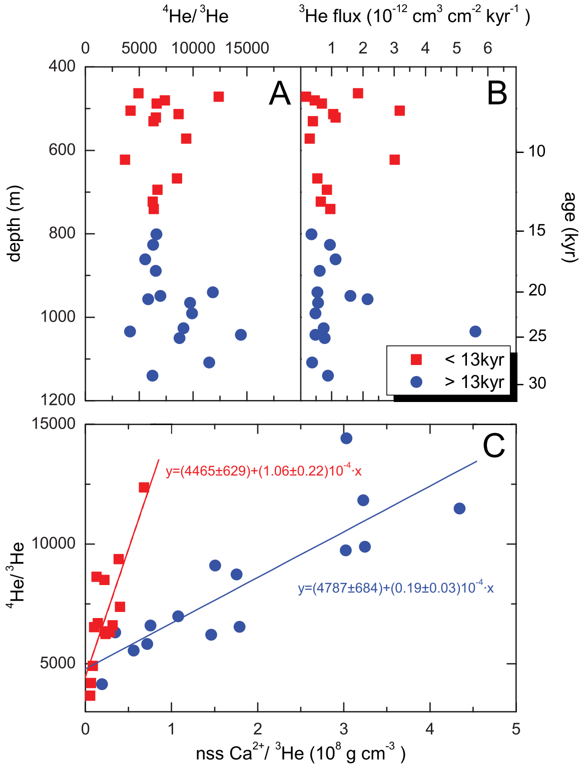


Fig. 1. Helium isotope characteristics of the ice samples from Dronning Maud Land (Antarctica). (**A** and **B**) The depth records of ^{4}He/^{3}He and the ^{3}He flux, respectively. Age scale is given on the right *y* axis. (**C**) An isotope mixing diagram with nss Ca^{2+} as the terrestrial dust reference species. Glacial and interglacial ice samples fall along two well-defined mixing lines with matching *y* intercepts (^{4}He/^{3}He ratio of the extraterrestrial end-member) and distinct slopes, suggesting a glacial-interglacial change in terrestrial dust source distribution.

Our data permit an independent estimate of the helium isotope ratio of the interplanetary dust deposited on Earth. An isotope mixing diagram (Fig. 1C) shows well-defined mixing lines with distinct terrestrial end-members, but anchored in the same extraterrestrial end-member, for both glacial and interglacial samples. The intercept with the ^{4}He/^{3}He axis, representing purely IDP-derived helium, is 4626 ± 465, indistinguishable from the average ^{4}He/^{3}He ratio of about 4170 ± 500 observed in studies of individual stratospheric IDPs (*7*).

The mixing diagram also indicates a change in the distribution of terrestrial dust sources between glacial and interglacial samples. Glacial ice shows ^{4}He/non–sea salt Ca^{2+} (nss Ca^{2+}) ratios that are much lower than those of the interglacial ice. This result is consistent with the moderate decrease of the terrigeneous ^{4}He flux from glacial to interglacial values by about a factor of 2, which is much lower than the 10- to 15-fold decrease observed in particulate dust flux measurements (*8*). We suggest that different dust sources, exposed continental shelves, or freshly generated glaciogenic material may have influenced the glacial dust deposition on the Antarctic ice sheet.

Our excess water technique enables parallel high-resolution reconstruction of extraterrestrial and terrestrial dust fluxes from ice cores. Large volume sampling, together with noble gas mass spectrometry, has opened a prospect to use IDPs and ^{3}He fluxes to pace the deposit of material and chemical signals in settings where the flow of time is otherwise only poorly constrained.

References and Notes

1. P. Gabrielli *et al.*, *Nature* **432**, 1011 (2004).
2. E. J. Brook, M. D. Kurz, J. Curtice, S. Cowburn, *Geophys. Res. Lett.* **27**, 3145 (2000).
3. Information on material and methods is available on *Science* Online.
4. G. Winckler, R. F. Anderson, M. Stute, P. Schlosser, *Quat. Sci. Rev.* **23**, 1873 (2004).
5. R. A. Muller, G. J. MacDonald, *Nature* **377**, 107 (1995).
6. F. Marcantonio *et al.*, *Paleoceanography* **16**, 260 (2001).
7. A. O. Nier, D. J. Schlutter, *Meteoritics* **27**, 166 (1992).
8. EPICA community members, *Nature* **429**, 623 (2004).
9. G.W. acknowledges support from the Comer Science and Education Foundation. We thank P. Schlosser, M. Stute, and R. F. Anderson for continued support of the lab. This work is a contribution to EPICA, a joint European Science Foundation/European Commission scientific program funded by the European Union and by national contributions from Belgium, Denmark, France, Germany, Italy, Netherlands, Norway, Sweden, Switzerland, and the UK. This is L-DEO publication 6901 and EPICA publication 154.

Supporting Online Material
www.sciencemag.org/cgi/content/full/313/5786/491/DC1
Materials and Methods

15 March 2006; accepted 16 May 2006
10.1126/science.1127469

[1]Lamont-Doherty Earth Observatory (L-DEO), Earth Institute at Columbia University, Palisades, NY 10964, USA. [2]Alfred Wegener Institute for Polar and Marine Research, Columbusstrasse, 27568 Bremerhaven, Germany.

*To whom correspondence should be addressed. E-mail: winckler@ldeo.columbia.edu

Plio-Pleistocene Ice Volume, Antarctic Climate, and the Global $\delta^{18}O$ Record

M. E. Raymo,[1]* L. E. Lisiecki,[1] Kerim H. Nisancioglu[2]

We propose that from ~3 to 1 million years ago, ice volume changes occurred in both the Northern and Southern Hemispheres, each controlled by local summer insolation. Because Earth's orbital precession is out of phase between hemispheres, 23,000-year changes in ice volume in each hemisphere cancel out in globally integrated proxies such as ocean $\delta^{18}O$ or sea level, leaving the in-phase obliquity (41,000 years) component of insolation to dominate those records. Only a modest ice mass change in Antarctica is required to effectively cancel out a much larger northern ice volume signal. At the mid-Pleistocene transition, we propose that marine-based ice sheet margins replaced terrestrial ice margins around the perimeter of East Antarctica, resulting in a shift to in-phase behavior of northern and southern ice sheets as well as the strengthening of 23,000-year cyclicity in the marine $\delta^{18}O$ record.

Although the glacial-interglacial cycles of the past 3 million years (My) represent some of the largest and most studied climate variations of the past, the physical mechanisms driving these cycles are not well understood. For the past 30 years, the prevalent theory has been that fluctuations in global ice volume are caused by variations in the amount of insolation received at critical latitudes and seasons because of variations in Earth's precession, obliquity, and eccentricity. Based mainly on climate proxy records from the past 0.5 My, but also supported by climate model results, a loose scientific consensus has emerged that variations in ice volume at precession [~23 thousand years (ky)] and obliquity (41 ky) frequencies appear to be directly forced and coherent with northern summer insolation, whereas the ~100-ky component of the ice age climate cycle results from nonlinear amplification mechanisms possibly phase-locked to summer insolation variations (*1*–*3*).

In the late Pliocene/early Pleistocene (LP/EP) interval from ~3 to 1 million years ago (Ma), however, only weak variance at 100-ky and 23-ky periods is observed in proxy ice volume records such as benthic $\delta^{18}O$. Instead, the records are dominated by 41-ky cyclicity, the primary obliquity period (Figs. 1A and 2A) (*4*–*6*) [supporting online material (SOM) text]. Given that the canonical Milankovitch model predicts that global ice volume is forced by high northern summer insolation, which at nearly all latitudes is dominated by the 23-ky precession period (Figs. 1B and 2A) (*7*), why then do we not observe a strong precession signal in LP/EP ice volume records? The lack of such a signal and the dominance of obliquity have defied understanding. Similarly, some ice modeling experiments show a dominant 41-ky periodicity, but there is always relatively more precession power in simulated ice volume than is observed in the geologic record; no ice sheet–climate model that we are aware of has been successful in reproducing the observed spectral characteristics of the LP/EP ice volume record (*8*–*10*). In every model, including our own recent ice modeling experiments that include meridional energy fluxes sensitive to varying insolation gradients (*5*, *10*), ablation is highly sensitive to summer heating and hence precession is always strongly represented in the predicted ice volume record. The strong influence of summer heating on ice sheet mass balance is also supported by more than a century of glaciological field studies [as summarized in (*11*) and shown in Fig. 3].

Here, we present a simple model of ice volume change, consistent with traditional Milankovitch theory and glaciological field studies, that predicts a sea level/$\delta^{18}O$ record that closely matches that observed from the geologic record. We used the nondimensional ice sheet–climate model of Imbrie and Imbrie (*12*), but a more sophisticated ice sheet model

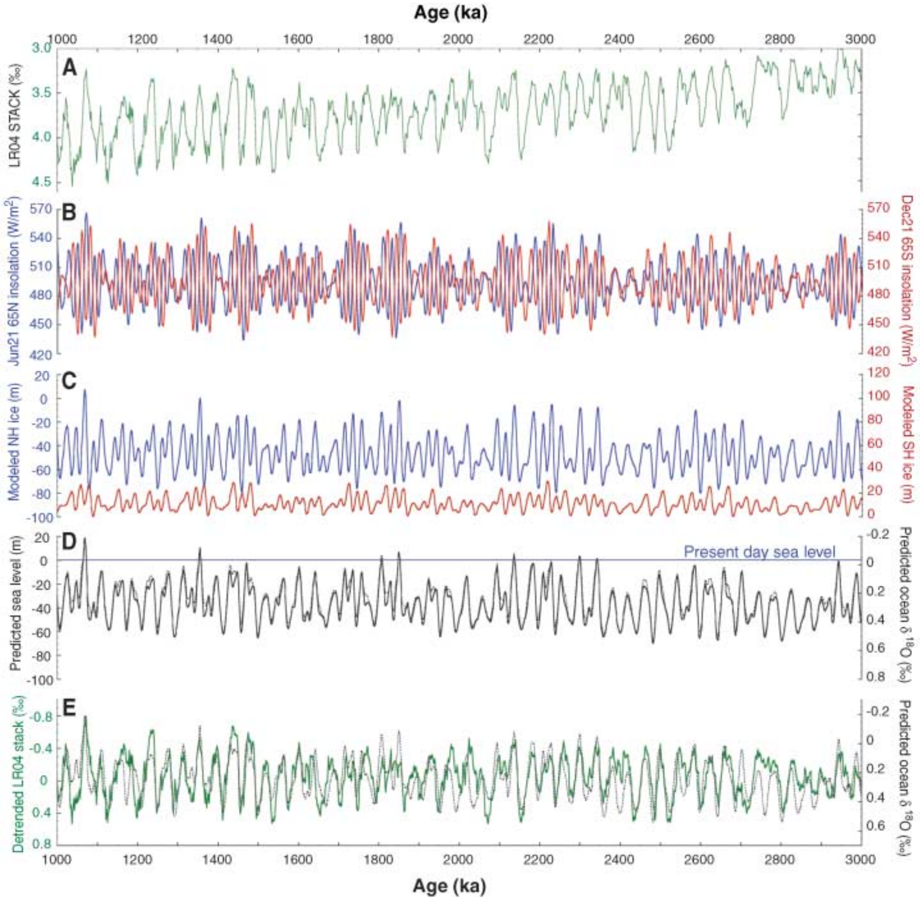


Fig. 1. Age versus (**A**) LR04 stack of >50 benthic $\delta^{18}O$ records (*6*); (**B**) 65°N summer insolation records for NH (21 June) and SH (21 December), calculated from (*7*); (**C**) NH (blue) and SH (red) modeled ice volumes, calculated as described in text; (**D**) predicted sea level (solid line) and mean ocean $\delta^{18}O$ (dashed line), derived from ice volume histories shown in (C); and (**E**) comparison of predicted mean ocean $\delta^{18}O$ and the LR04 stack detrended by a slope of 0.8‰ per My from 3 to 2.5 Ma and 0.26‰ per My from 2.5 to 1 Ma (*31*).

[1] Department of Earth Science, Boston University, 685 Commonwealth Avenue, Boston, MA 02215, USA. [2] Palaeoclimates, Bjerknes Center for Climate Research, Allegaten 55, Bergen 5007, Norway.

*To whom correspondence should be addressed. E-mail: raymo@bu.edu

would give similar results (*10*) (SOM text). Our modeled ice sheets are dominated by precession because of the assumed (and observed) dependence of ablation on summer temperatures. Our experiment differs from previous attempts to model the "41-ky world" because we allowed for a dynamic Antarctic ice sheet, as suggested by Pliocene sea level data. First, we present evidence for a more dynamic Antarctic ice sheet in the LP/EP, followed by model results and a discussion of the implications of our hypothesis.

Mid-Pliocene climate and ice sheet margins. Many marine and terrestrial studies have documented the long-term cooling that began in the early Pliocene and culminated in the growth of large Northern Hemisphere (NH) ice sheets by 2.5 Ma (*4*, *13–15*). It is also widely recognized that the mid-Pliocene before 2.9 Ma was the most recent time period consistently warmer than the present, with global temperatures elevated by as much as 3°C with respect to modern values (*16*). In particular, the interval between 3.3 and 3.0 Ma, often referred to as the ''mid-Pliocene climatic optimum,'' is widely studied as a possible analog for a future warmer Earth (*17*).

From 3.3 to 3.0 Ma, the deep ocean $\delta^{18}O$ record is characterized by consistently more depleted isotopic values (lower than modern values by >0.5‰), indicative of warmer bottom waters and/or less global ice volume (*14–17*). Independent evidence for higher sea levels during the mid-Pliocene climate optimum comes from raised coastal terraces [35 ± 18 m relative to present (rtp) (*18*)] and Pacific atolls [up to 25 m higher rtp (*19*)]. Given that the present-day Greenland and West Antarctic ice sheet volumes are each equivalent to only 6 to 7 m of sea level (*20*), the above studies imply that a substantial volume of the present East Antarctica ice sheet (EAIS) must have melted at this time [today the EAIS is equivalent to ~54 to 55 m of sea level (*21*)]. Studies conducted on and around Antarctica suggest a warmer, partially deglaciated EAIS at this time, including extensive paleosol development (*22*), increased smectite in near-shore sediment (*23*), and less regional ice-rafted material (*24*). Recent expeditions have also found evidence for dynamic behavior of the EAIS margin throughout the Plio-Pleistocene, including less continental ice, reduced sea-ice cover, and inland penetration of warmth in the Prydz Bay region (*25*, *26*), as well as a substantial melting of the Ross ice shelf, near 1.0 Ma (*27*). Indeed, the almost completely ice-covered and poorly studied EAIS coastline (generally located between 65°S and 70°S over more than 7000 km) could have been deglaciated (melting ice sheet margin on land) for much of the late Pliocene and/or early Pleistocene, leaving little evidence today.

Ultimately, ice sheets are at the mercy of the competing forces of ablation and accumulation (Fig. 3). In East Antarctica today (Fig. 3B), virtually no melting occurs and precipitation is limited by low air temperature. Most ablation is due to calving of icebergs from ice margins at sea level (*28*) (on the West Antarctic Peninsula, by contrast, summer temperatures exceed 0°C and grass and mosses take root today). During the last glacial maximum (21 ka), Antarctica is believed to have increased in volume by 15 m of sea level equivalent (*29*), most likely by expanding onto exposed shelves as sea level fell because of NH ice sheet growth (note, in Fig. 3, that glacial cooling in and of itself would predict a decrease in mass accumulation). By comparison, Greenland today (hatched bar in Fig. 3B) experiences widespread summer melting in low altitude coastal regions that is offset by accumulation inland. During the last glacial maximum, the expanded Laurentide and Fennoscandia ice sheets would also have experienced widespread summer melting on their southern margins.

From modern glaciological observations and paleo–sea level data, we draw this conclusion: The deglaciation of a substantial fraction of the EAIS at 3 Ma, suggests that the EAIS behaved glaciologically, at that time, like a modern Greenland ice sheet. In other words, the EAIS must have overlapped the range of negative mass balance (uppermost bar in Fig. 3B). A warmer, more dynamic EAIS with a terrestrial-based melting margin, as opposed to a glaciomarine calving margin, is implied. Because such margins are strongly controlled by summer melting, Antarctic ice volume would be sensitive to orbitally driven changes in local summer insolation. When did the EAIS transition to its modern state, ringed by extensive marine ice shelves? Until now it has been assumed that it happened in concert with the well-documented NH cooling between 3 and 2.6 Ma. Here, we propose that it may not have happened until after 1 Ma.

Modeled Plio-Pleistocene ice volume history. Next, we present a forward model of global ice volume history initialized at 3 Ma with the following assumptions: (i) ice sheet mass balance is sensitive to local summer insolation; (ii) NH ice volume varies on orbital time scales between the present volume and 80 m below present sea level; and (iii) Antarctic ice volume varies between the present value and sea level that is 30 m higher than the present sea level. In other words, cool NH summers will lead to NH ice growth while, at the

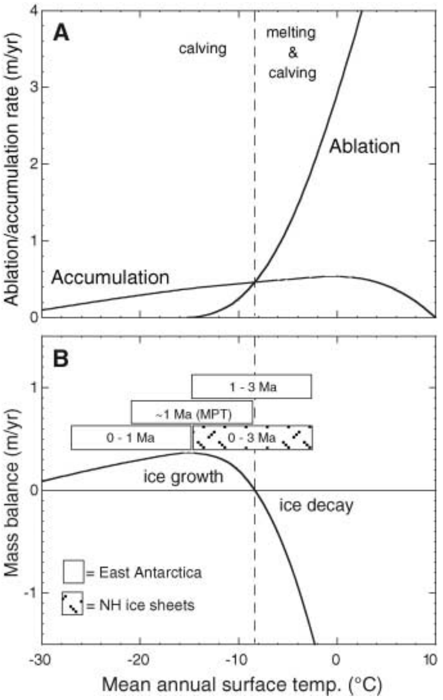


Fig. 3. Generalized dependence of ice sheet ablation rate, accumulation rate (**A**), and mass balance (**B**) on mean annual surface temperature [modified from (*58*)]. Hatched and open bars show hypothesized time evolution of NH and SH ice sheets, respectively. Vertical dashed line denotes transition from ablation-dominated to accumulation-dominated regime. MPT, mid-Pleistocene transition.

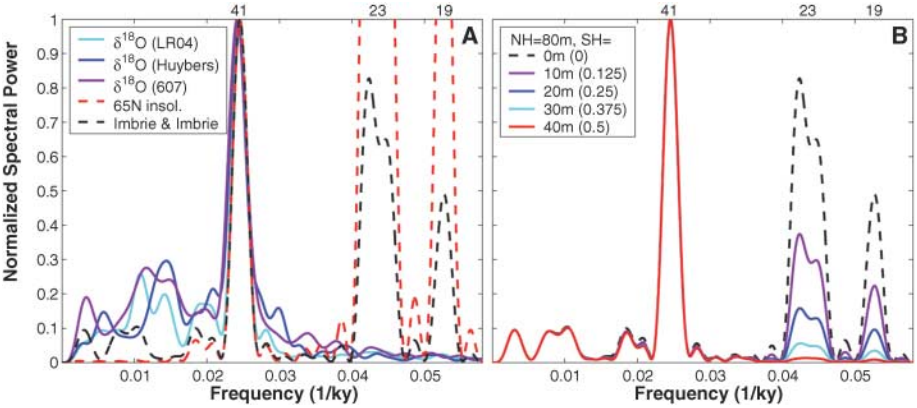


Fig. 2. (**A**) Spectra of the LR04 stack, the paleomagnetically dated $\delta^{18}O$ stack of Huybers (*57*), the paleomagnetically dated DSDP607 benthic $\delta^{18}O$ record (*4*, *34*), 21 June summer insolation at 65°N, and NH model output (Fig. 1C). All spectra are calculated over interval from 3 to 1 Ma and are normalized to each other (SOM text). (**B**) Comparison of spectra for sea level curves calculated with the use of different ratios of NH to SH ice volume change (ratio SH/NH ice given in parentheses). NH ice is always assumed to range over 80 m sea level equivalent and SH ice varies over a range of 0 to 40 m. In Fig. 1 we show the results of the 80 m/30 m experiment.

same time, warm Southern Hemisphere (SH) summers lead to ice decay in Antarctica (Fig. 1B). To predict the individual ice volume histories for each hemisphere, we used the well-known ice-climate model of Imbrie and Imbrie (*12*):

$$-dV/dt = (i + V)/\tau \qquad (1)$$

where V is ice volume, i is insolation (21 June, 65°N for NH; 21 December, 65°S for SH), and τ is a time constant which differs for ice growth and decay (see SOM text for more model details). Insolation and the modeled ice volume histories for the NH and SH (in sea level equivalents) are shown in Fig. 1, B and C. Individual ice sheet histories are dominated by both precession and obliquity frequencies (Fig. 2B), as would be expected.

Combining the two modeled ice sheet histories, one can predict global sea level (Fig. 1D). In the global ice volume/sea level signal, precession-driven responses, which are out of phase between the hemispheres, largely cancel each other out, leaving a record dominated by obliquity (Figs. 1D and 2B). Similar results would be found for any comparable ratio of northern to southern ice volume. The above assumptions about ice sheet evolution are simplistic; for instance, ice-rafted detritus (IRD) records suggest that NH ice sheets were increasing in volume over the interval from 2.9 to 2.5 Ma (*4*, *13*), whereas we assume no long-term volumetric trends in the ice sheets. If one allowed NH ice volume to gradually increase from 3 to 1 Ma, one would expect to observe a gradual increase in precession power in modeled sea level. Such an increase is observed in δ^{18}O data (SOM text and fig. S1).

One can convert modeled ice volume to δ^{18}O units by making an assumption about the mean δ^{18}O of ice at each pole (*30*). We then compared the predicted mean ocean δ^{18}O to the LR04 δ^{18}O stack (*6*) after detrending the stack for long-term global cooling (Fig. 1E). Despite some obvious mismatches in amplitude and/or structure, the overall correspondence between our model output and a global stack of more than two dozen benthic δ^{18}O records is excellent (*31*). The ability of this simple model to recreate the "41-ky world" suggests our hypothesis, the partially out-of-phase waxing and waning of ice sheets in both hemispheres over much of the Plio-Pleistocene, merits consideration. The data and model mismatches may also arise from the temperature component, time-scale errors, and geologic noise contained in the LR04 stack. Our conclusion is relatively insensitive to the sea level ranges and/or isotopic compositions assumed (sensitivity tests shown in Fig. 2B) or values chosen for the time constants in the model (SOM text). A ratio of SH to NH ice of just 13% (10 m SH/80 m NH) results in a pronounced diminishment of the precessional signal in the modeled δ^{18}O/sea level record, and a ratio of 25% (20 m SH/80 m NH) results in the appearance of a "41-ky world."

Other climate proxy records. The above model reconciles the LP/EP δ^{18}O record with evidence drawn from modern glaciological studies, ice sheet–climate models, and recent ice sheet history for the strong control exerted by summer temperatures on ablation. Our hypothesis is also consistent with the presence of large ice sheets in the mid-latitudes of the United States in the LP/EP (*9*), as well as with an inferred 23-ky periodicity in melt water delivery down the Mississippi River drainage at that time (*32*). One might argue that it would require an unrealistically large warming to develop a terrestrial melting margin on the EAIS. Yet sediments recovered from an ice-covered lake in the Prydz Bay area show the presence of running water, warmer water diatoms, and mosses during the penultimate interglaciation (*33*), widely recognized as being only slightly warmer than the Holocene (*20*).

More difficult to reconcile with our proposed NH and SH ice volume histories are proxy records of sea surface temperature (SST) and IRD from the Northern and Southern Atlantic Ocean that show a strong 41-ky pacing (*4*, *13*, *24*, *34*, *35*). Indeed, the covariance of the δ^{18}O, IRD, and SST records in the high latitude North Atlantic has long been invoked as sedimentological evidence that the variability observed in benthic δ^{18}O must derive in large part from the waxing and waning of ice sheets at the 41-ky periodicity in the NH (*4*, *13*, *34*). How then could large ice volume changes at the precessional period be missed? For the IRD record, we propose that the answer lies in the behavior of the two types of ice sheet margins: terrestrial and glaciomarine. On a terrestrial margin, ice sheet advance and retreat is strongly controlled by surface melt that is almost entirely dependent on summer heating. Such margins leave no imprint on marine IRD records because they are not in contact with the ocean. On the other hand, glaciomarine margins, similar to more than 90% of the Antarctic ice margin today, are the source of icebergs that deliver IRD to open ocean. Such margins are highly sensitive to sea level variations that can unpin and destabilize ice margins grounded below sea level (*28*). Indeed, both the early and late Pleistocene records of IRD in the North Atlantic show the most notable input occurring on deglaciations during which sea level is rising the fastest (*35*–*37*). In summary, calving rates on marine-based margins are controlled primarily by sea level and hence would be expected to follow the 41-ky sea level record (*38*).

The SST signal of the high-latitude Atlantic has also been shown to vary primarily at 41-ky between 1.6 and 1 Ma (*34*) (before this time, SST estimates are problematic because of no-analog/extinct species). In the late Pleistocene, Atlantic SST varies at both precession and obliquity periods; however, the obliquity rhythm dominates at latitudes of >50°N, where negligible precession is observed (*39*). At latitudes of <50°N, precession dominates with obliquity essentially disappearing south of 40°N (*39*). The controls of SST in the North Atlantic are poorly understood, although clearly late Pleistocene SST records poleward of 50°N are dominated almost exclusively by obliquity despite the known presence (from coral reef records) of 23-ky variability in ice volume (*40*). It may be that large changes in the extent of winter sea ice, possibly sensitive to mean annual or winter insolation at high latitudes (obliquity controlled), exert a more direct influence on polar and subpolar SST (*11*, *41*).

Beyond the North Atlantic region, numerous proxy records are dominated by precession, obliquity, or both frequencies in the LP/EP. None of these records rules out the existence of a precessional signal in NH or SH ice volume. African dust records (*42*) and grain-size variations in Chinese loess records (*43*) exhibit both precession and obliquity variance throughout the LP/EP. Climate-sensitive proxies from the Mediterranean region also show precession and obliquity pacing throughout the past 3 My (*44*). By contrast, tropical Pacific records (*45*, *46*) show an almost exclusive 41-ky signal in SST, although it leads δ^{18}O and hence cannot be responding to ice sheet forcing. These latter proxies are sensitive to the strength of trade winds and/or westerlies, which in turn are sensitive to meridional insolation gradients and thus obliquity (*5*, *10*, *47*).

Mid-Pleistocene transition. We know from ice core and coral reef records that late Pleistocene temperature and ice volume variations are roughly in phase between both hemispheres (*29*, *48*) and that sea level variations were paced by NH summer insolation forcing (*40*). We argue that this pattern of climate change was the inevitable consequence of long-term cooling that gradually drove the EAIS margin into the sea. We suggest that by ~1.0 Ma, high-latitude climate had cooled to the extent that it was no longer warm enough for an extensive terrestrial melting margin to exist on East Antarctica (middle bar in Fig. 3B). Ablation now occurred primarily by means of calving, and accumulation over the entire ice sheet may have resulted in the progressive thickening of the EAIS, limited only by ice stream drawdown mechanisms and moisture starvation.

Implicit in this scenario is the conclusion that sea level changes driven by NH ice sheet fluctuations became the primary control on Antarctic ice volume after ~1 Ma. When sea level dropped, the EAIS would grow out onto the continental shelf; when sea level rose, the retreat of the marine ice sheet grounding lines around Antarctica would result in rapid ice shelf disintegration. Ice volume at both poles would now vary in phase at both obliquity and precession frequencies, and δ^{18}O would thus exhibit both 23-ky and 41-ky cyclicity (as observed). These two modes of SH response (in phase versus out of phase) do not necessarily require an abrupt transition.

Conclusions. By allowing modest variations in Antarctic ice sheet size from 3 to 1 Ma, controlled by local insolation, we show that the dominant 41-ky period in marine $\delta^{18}O$ records may result from out-of-phase ice sheet growth at each pole. Individual ice volume histories in the Arctic and Antarctic realm were likely dominated by both precession (out of phase between poles) and obliquity (in phase between poles) with ice ablation strongly controlled by summer temperatures. Our hypothesis solves the conundrum of why no strong precession signal is observed in global $\delta^{18}O$ records from this time despite the well-known importance of summer temperatures on ice sheet and glacier mass balance (*49*). Our hypothesis also predicts the presence of a dynamic EAIS in the LP/EP characterized by a terrestrial ablation margin at latitudes between 65°S and 70°S. We also predict that the record of local temperature recorded by deuterium isotopes in ice cores (should ice this old ever be recovered) would be in phase with SH insolation at the precession frequency. In the NH, sites sensitive to the southern margin of the NH ice sheet should show a record of variability much like that depicted in Fig. 1C.

We further propose that long-term cooling resulted in a transition from a primarily land-based to primarily marine-based EAIS margin about 1.0 Ma, resulting in the mid-Pleistocene transition and the strengthening of 23-ky cycles in the $\delta^{18}O$ record. Ice sheet volume may have increased at both poles at this time because of the establishment of positive globally synchronous feedbacks (such as albedo and CO_2) at the precession frequency (*50*). Lastly, the strengthening of CO_2 and albedo feedbacks by enhanced sea level fall or aridity, in conjunction with long-term global cooling, may have led to the establishment of NH ice sheets large enough to survive summer insolation maxima of low intensity, a necessary prerequisite for the development of the ‘‘100-ky’’ cycle (*51*).

References and Notes

1. J. Imbrie *et al.*, *Paleoceanography* **7**, 701 (1992).
2. D. Paillard, *Rev. Geophys.* **39**, 325 (2001).
3. E. Tziperman, M. E. Raymo, P. Huybers, C. Wunsch, *Paleoceanography*, in press.
4. M. E. Raymo, W. F. Ruddiman, J. Backman, B. M. Clement, D. G. Martinson, *Paleoceanography* **4**, 413 (1989).
5. M. E. Raymo, K. Nisancioglu, *Paleoceanography* **18**, 10.1029/2002PA000791 (2003).
6. L. E. Lisiecki, M. E. Raymo, *Paleoceanography* **20**, 10.1029/2004PA001071 (2005).
7. J. Laskar, F. Joutel, F. Boudin, *Astron. Astrophys.* **270**, 522 (1993).
8. A. Berger, X. S. Li, M. F. Loutre, *Quat. Sci. Rev.* **18**, 1 (1999).
9. P. U. Clark, D. Pollard, *Paleoceanography* **13**, 1 (1998).
10. K. H. Nisancioglu, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA (2004); available online (http://hdl.handle.net/1721.1/16703).
11. G. H. Denton, R. B. Alley, G. Comer, W. S. Broecker, *Quat. Sci. Rev.* **24**, 1159 (2005).
12. J. Imbrie, J. Z. Imbrie, *Science* **207**, 943 (1980).
13. N. J. Shackleton *et al.*, *Nature* **307**, 620 (1984).
14. M. E. Raymo, *Annu. Rev. Earth Planet. Sci.* **22**, 353 (1994).
15. M. Mudelsee, M. E. Raymo, *Paleoceanography* **20**, 10.1029/2005PA001153 (2005).
16. A. C. Ravelo, D. H. Andreasen, M. Lyle, A. O. Lyle, M. W. Wara, *Nature* **429**, 263 (2004).
17. H. J. Dowsett *et al.*, *Global Planet. Change* **9**, 169 (1994).
18. H. J. Dowsett, T. M. Cronin, *Geology* **18**, 435 (1990).
19. B. R. Wardlaw, T. M. Quinn, *Quat. Sci. Rev.* **10**, 247 (1991).
20. K. M. Cuffey, S. J. Marshall, *Nature* **404**, 591 (2000).
21. P. Huybrechts, D. Steinhage, F. Wilhelms, J. L. Bamber, *Ann. Glaciol.* **30**, 52 (2000).
22. G. J. Retallack, E. S. Krull, J. G. Bockheim, *J. Geol. Soc. London* **158**, 925 (2001).
23. J. Junttila, K. Strand, *Eos Trans. AGU* **86**, Fall Meet. Suppl., abstract PP41B-0639 (2005).
24. L. Murphy, D. A. Warnke, C. Andersson, J. Channell, J. Stoner, *Paleoceanography* **182**, 183 (2002).
25. P. G. Quilty, *Eos Trans. AGU* **86**, Fall Meet. Suppl., abstract PP51F-04 (2005).
26. A. K. Cooper, P. E. O'Brien, in *Proc. Ocean Drill. Prog. Sci.*, A. K. Cooper, P. F. O'Brian, C. Richter, Eds. (2004), vol. 188, ch. 1; available online (http://www-odp.tamu.edu/publication/188_SR/synth/synth.htm).
27. R. P. Scherer, *Eos Trans. AGU* **86**, Fall Meet. Suppl., abstract PP43C-05 (2005).
28. D. I. Benn, D. J. A. Evans, *Glaciers and Glaciation* (Arnold Press, London, 1998).
29. S. E. Bassett, G. Milne, J. X. Mitrovica, P. U. Clark, *Science* **309**, 925 (2005).
30. Assuming NH ice averages –30 per mil (‰) and SH ice averages –45‰ results in a predicted ocean $\delta^{18}O$ amplitude that is a little more than half of that observed (implying deep ocean temperature changes of a few degrees; Fig. 1, D and E). The difference in NH and SH ice sheet isotopic composition also results in a slight decrease of the precessional component of the modeled curve (dashed line in Fig. 1D). If NH glacial ice is assumed to be more depleted, –40‰ for example, then the modeled $\delta^{18}O$ approaches the amplitude of the LR04 stack (Fig. 1A), implying that much of LP/EP signal is due to ice volume change. Ultimately, ongoing development of independent temperature proxies may allow us to isolate the ice volume component in ocean $\delta^{18}O$ records and provide a better modeling target.
31. The overall correlation coefficient between stack and modeled $\delta^{18}O$ from 2.7 to 1.2 Ma is 0.68. Small age model adjustments to the stack, within the range of age model uncertainty (<5 ky), can increase the correlation to 0.75. Both correlation values decrease slightly (to 0.63 and 0.72, respectively) when the full interval of 3 to 1 Ma is used. The stack is detrended by a slope of 0.8‰ per My from 3 to 2.5 Ma and 0.26‰ per My from 2.5 to 1 Ma.
32. J. E. Joyce, L. Tjalsma, J. Prutzman, *Geology* **21**, 483 (1993).
33. D. A. Hodgson *et al.*, *Quat. Sci. Rev.* **25**, 179 (2006).
34. W. F. Ruddiman, M. E. Raymo, D. G. Martinson, B. M. Clement, J. Backman, *Paleoceanography* **4**, 353 (1989).
35. K. A. Venz, D. A. Hodell, C. Stanton, D. A. Warnke, *Paleoceanography* **14**, 42 (1999).
36. M. E. Raymo, K. Ganley, S. Carter, D. W. Oppo, J. McManus, *Nature* **392**, 699 (1998).
37. J. F. McManus, D. W. Oppo, J. L. Cullen, *Science* **283**, 971 (1999).
38. High-resolution studies (*36*, *37*) also reveal that episodes of less notable IRD input occur on millennial time scales (Dansgaard-Oeschger/Heinrich events), events that are believed to reflect solar or stochastically forced instability in the marine ice margin grounded in Hudson Bay [reviewed in (*52*)]. It may be that successive increments of sea level fall, obliquity-paced in our hypothesis, require the marine-based components of ice sheets to constantly readjust their grounding lines seaward. Millennial-scale interruptions in the delivery of IRD to the open ocean may occur when an ice margin is growing out to a new stable grounding line. When a new grounding line is established, the renewed calving of icebergs would occur, contributing to regional cooling and freshening of ocean surface waters and thus promoting the development of sea ice cover and weakening thermohaline convection. In other words, one need not invoke any millennial-scale forcing but only the slow relentless drive of orbital-scale changes in sea level.
39. W. F. Ruddiman, A. McIntyre, *Geol. Soc. Am. Bull.* **95**, 381 (1984).
40. W. G. Thompson, S. L. Goldstein, *Quat. Sci. Rev.,* in press.
41. M. E. Raymo, D. Rind, W. F. Ruddiman, *Paleoceanography* **5**, 367 (1990).
42. P. B. deMenocal, *Science* **270**, 53 (1995).
43. Y. Sun, S. C. Clemens, Z. An, Z. Yu, *Quat. Sci. Rev.* **25**, 33 (2006).
44. L. J. Lourens, F. J. Hilgen, J. Gudjonsson, W. J. Zachariasse, *Mar. Micropaleontol.* **19**, 49 (1992).
45. Z. Liu, T. D. Herbert, *Nature* **427**, 720 (2004).
46. M. Medina-Elizalde, D. W. Lea, *Science* **310**, 1009 (2005).
47. S.-Y. Lee, C. J. Poulsen, *Paleoceanography* **20**, 10.1029/2005PA001161 (2005).
48. T. Blunier, E. J. Brook, *Science* **291**, 109 (2001).
49. By contrast, in a late Oligocene/early Miocene benthic $\delta^{18}O$ record (*53*), both precession and obliquity variance are observed. This result is consistent with the presence of ice sheets in only one hemisphere (Antarctica) at that time.
50. In the late Pleistocene, ice core CO_2 records are strongly paced by obliquity and more weakly influenced by precession (*54*). Given the uncertainties in the controls of atmospheric CO_2, our hypothesis cannot predict how the early Pleistocene CO_2 record will vary. For instance, if CO_2 variations are controlled by sea level (*55*), then the development of in-phase polar climate behavior at the mid-Pleistocene transition would cause global CO_2 variations to occur at the precession period for the first time. If high-latitude aridity and dust supply control CO_2 (*56*), then the possible dominance of one hemisphere over the other (due to greater land mass for instance) could impart a precession signal to the early Pleistocene CO_2 record. In this case, we might still expect the precession signal in the CO_2 record to get stronger at the mid-Pleistocene transition if high-latitude aridity, as well as ice volume, began to vary in phase. Lastly, if positive climate feedbacks act preferentially on obliquity time scales (*54*), then the ice volume signal at the 41-ky period could be additionally amplified relative to precession.
51. M. E. Raymo, *Paleoceanography* **12**, 577 (1997).
52. G. C. Bond *et al.*, in *Geophys. Monogr. Am. Geophys. Union 112* (American Geophysical Union, Washington, DC, 1999), pp. 35–58.
53. J. C. Zachos, N. J. Shackleton, J. S. Revenaugh, H. Pälike, B. P. Flower, *Science* **292**, 274 (2001).
54. W. F. Ruddiman, *Quat. Sci. Rev.* **22**, 1597 (2003).
55. J. C. Latimer, G. M. Filippelli, *Paleoceanography* **16**, 627 (2001).
56. R. B. Alley, E. J. Brook, S. Anandakrishnan, *Quat. Sci. Rev.* **21**, 431 (2002).
57. P. Huybers, *Quat. Sci. Rev.*, in press.
58. P. Huybrechts, in *Climate of the 21st Century: Changes and Risks*, J. Lozan, H. Grabl, P. Hupfer, Eds. (GEO Wissenschaftliche Auswertungen, Hamburg, 2001), pp. 221–226.
59. We thank K. Lawrence, J. Fastook, D. Marchant, B. Ruddiman, P. Huybrechts, J. Kennett, B. Curry, R. Scherer, D. Bowen, P. Huybers, and D. Oppo for assisting us on this project, either by reading the manuscript or providing helpful guidance and information; E. Tziperman for ongoing discussions of Milankovitch and climate; and four anonymous reviewers whose comments greatly improved the manuscript. M.E.R. acknowledges the support of NSF grant ATM-0220681. L.E.L. is supported by a NOAA Climate and Global Change postdoctoral fellowship.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1123296/DC1
SOM Text
Figs. S1 to S4
Tables S1 to S3
References

1 December 2005; accepted 9 June 2006
Published online 22 June 2006;
10.1126/science.1123296
Include this information when citing this paper.

Transcriptional Repression Distinguishes Somatic from Germ Cell Lineages in a Plant

Farzad Haerizadeh, Mohan B. Singh, Prem L. Bhalla*

In flowering plants, the male germline begins with an asymmetric division, after which one of the resulting cells, the generative cell, divides symmetrically to produce two sperm cells. We show here that the male germline is initiated by transcriptional control. We identify GRSF, germline-restrictive silencing factor, from the lily. GRSF is ubiquitous in nongerm cells and is absent from male germ cells. GRSF recognizes silencer sequences in promoters of genes specific to the germline, stably repressing these genes in cells that are not destined to become germ cells.

Sexual reproduction in flowering plants requires a pair of sperm that travel together through the pollen tube to the embryo sac. The pair of sperm results from symmetric cell division of the generative cell, which is in turn the result of a preceding asymmetric cell division. Unlike in animals, where the male germline is set aside early in development, in plants the male germ lineage arises from cells of a previously somatic lineage. Some of the gene expression patterns operating in these male germline cells have been identified (*1*, *2*). The genes and proteins that are essential for the unique functions of male germline cells in fertilization are likely to be among those expressed specifically in these cells. Transcripts have been identified that are expressed only in sperm cells or their precursor generative cells (*3*–*6*). The gene *LILY GENERATIVE CELL-SPECIFIC 1* (*LGC1*) is a male germline–specific gene (*3*, *7*), and its promoter, which contains a silencer region (*7*), can direct the expression of reporter genes in male germline cells of transgenic plants (*7*). We used the *LGC1* gene promoter to study the regulatory mechanisms that control developmental gene expression in the male germline of *Lilium* (lily) and *Arabidopsis*. We found that an essential component of male germline cell–specific regulation of *LGC1* and other coordinately expressed genes lies in a germline restrictive silencing factor (GRSF) that represses their expression in other plant cells.

We prepared a *LGC1* promoter–red fluorescent protein (p*LGC1-RFP*) construct and confirmed its male germline specificity and the presence of a silencer region (fig. S1, A to C). We further reasoned that if in non–male germline cells the *LGC1* promoter is repressed by the binding of a specific repressor to the silencer element, then flooding these cells with an excess of silencer sequence should lead to derepression of the promoter. We designed a competitor comprising 16 ligated repeats of 43–base pair (bp) double-stranded oligonucleotide silencer sequences (fig. S1D) and tested it in a transient expression system that involved electroporation-mediated cotransformation of the competitor and p*LGC1-RFP* into lily petal protoplasts (fig. S1E). In the absence of competitor, no expression of RFP was observed. However, cotransformation with competitor led to a reactivation of *LGC1* promoter. We further noticed that increasing the ratio of the competitor versus the reporter construct concomitantly enhanced the level of expression of RFP (Fig. 1A). A p*CaMV35S-EGFP* reporter construct (EGFP, enhanced green fluorescent protein) was also cotransformed in all experiments as an internal control of electroporation efficiency. The presence or absence of competitor had no effect on the expression of GFP under the control of *CaMV35S* promoter. These results revealed that sequestration of repressor by the excess silencer sequences can lead to ectopic activation of *LGC1* promoter. Such competitor-induced ectopic activation of *LGC1* promoter was also observed when p*LGC1-RFP* construct was introduced in lily petal cells by microprojectile bombardment (Fig. 1B and fig. S1F). These data suggested that the in vivo repression of *LGC1* promoter in non–male germline cells might be mediated by the binding of a sequence-specific repressor.

***LGC1* silencer sequence can recruit silencing machinery to a heterologous promoter.** To investigate whether the *LGC1* silencer sequence is sufficient to recruit transcriptional silencing machinery in vivo, we replaced a 43-bp sequence from a constitutive (*CaMV35S*) promoter (base pairs –216 to –259) with a 43-bp nucleotide sequence from *LGC1* promoter (Fig. 2A). This modified promoter fused with the *EGFP* coding sequence (p*mCaMV-EGFP*) (Fig. 2B) was cotransformed with p*CaMV-RFP* construct into lily petals. As shown in Fig. 2C, RFP signals but not GFP signals were observed in petal cells, indicating that the introduction of silencer sequences from *LGC1* promoter leads to complete inactivation of *CaMV* promoter. However, this modified *CaMV* promoter could be reactivated by cotransformation with an excess of the competitor (Fig. 2D). This activation is attributable to a lack of repressor binding to the silencer region in the modified *CaMV* promoter due to sequestration of repressor by the competitor. This silencing of *CaMV* promoter is sequence specific, because replacement of the same *CaMV* promoter domain with a randomly selected human genome sequence (Fig. 2E) had no noticeable effect on the promoter activity (Fig. 2F). These data show that the silencer sequence is sufficient to recruit a specific repressor and associated transcriptional silencing machinery in the context of a constitutive heterologous promoter.

Repressor recognition of similar silencer sequences is conserved in flowering plants. Lily *LGC1* promoter retains its strict generative and sperm cell specificity in a taxonomically distant

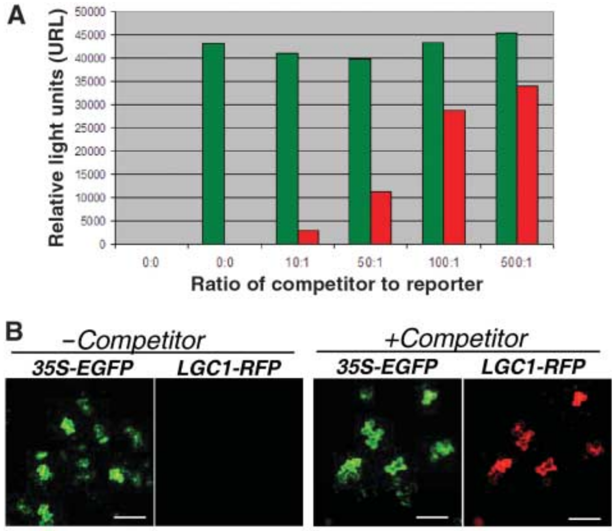


Fig. 1. Transcriptional activity of *LGC1* promoter. (**A**) Quantitative fluorometric assay showed that *LGC1* promoter is inactive in lily petal protoplasts, but cotransformation with increasing concentrations of competitor led to concomitant increases in the level of RFP expressed as a result of activation of *LGC1* promoter. In all electroporation experiments, *CaMV35S-EGFP* was cotransformed as an internal control. The amounts of RFP and GFP in biological replicates were quantified with a FLUOstar Optima microplate reader. The plate reader did not detect any signal in electroporated protoplast without the reporter constructs. Green and red bars denote GFP and RFP levels, respectively. (**B**) Transformation of lily petal cells with mixture of *LGC1-RFP*, *CaMV35S-EGFP*, and competitor led to activation of both *LGC1* and *CaMV35S* promoters, as reflected by the expression of RFP and GFP in the same cell. The expression of GFP as a result of cotransformation with *CaMV35S-EGFP* in all bombardment experiments acted as a positive control. Scale bars, 200 μm.

Plant Molecular Biology and Biotechnology Laboratory, Australian Research Council Centre of Excellence for Integrative Legume Research, Faculty of Land and Food Resources, University of Melbourne, Parkville, Victoria 3010, Australia.

*To whom correspondence should be addressed. E-mail: premlb@unimelb.edu.au

plant, tobacco (*7*), implying the conservation of a related sequence-specific repressor common to disparate families of flowering plants. Support for a conserved regulatory mechanism is extended by our further bombardment experiments. We cotransformed petal tissues from taxonomically diverse plants with p*delLGC1-EGFP* constructs, with p*CaMV-RFP* used as an internal control. The *delLGC1* promoter carrying a deletion of the silencer region led to constitutive expression of GFP in petal tissues from all tested plants. These results suggest the presence of functionally conserved repressors in other plants. Further experiments involving cotransformation with p*LGC1-RFP* and silencer competitor led to activation of the *LGC1* promoter in petal cells from such diverse plants as *Brassica*, *Magnolia*, and pea (Fig. 2G), providing evidence for the presence of an evolutionarily conserved repressor in flowering plants.

Repressor-mediated transcriptional regulation is functional in planta. *Arabidopsis* plants carrying *LGC1:GUS* construct showed no β-glucuronidase (GUS) activity (fig. S2A) relative to constitutive expression in plants carrying *del-LGC1-GUS* construct (fig. S2B). Thus, stable transformation provides accompanying in planta evidence that the silencer sequence in the *LGC1* promoter is essential for repression in non–male germline cells. Furthermore, the introduction of silencer sequences from *LGC1* promoter completely repressed *CaMV* promoter (fig. S2, C and D), hence this silencer can confer in planta transcriptional repression in the context of a surrogate promoter. However, *CaMV* promoter carrying randomly selected human genome sequence in the same location showed normal GUS expression in *Arabidopsis* plants (fig. S2E). These results provide confirmation of an evolutionarily conserved repressor system that is capable of regulating genes containing a *LGC1*-type cis-acting silencer sequence.

Germline restrictive silencing factor (GRSF) is a 24-kD protein. We identified the putative repressor, GRSF, by screening a lily petal cDNA expression library with double-stranded 43-bp radiolabeled silencer oligonucleotide, using in vitro binding conditions optimized by electrophoretic mobility shift assay (EMSA) with lily petal nuclear extracts (fig. S3). Screening of nearly 800,000 clones from an unamplified cDNA library led to the selection of four positive clones, the sequencing of which revealed that all four represented the same cDNA of varying lengths. One of these clones contained an 840-bp insert, and we used this sequence to obtain the full-length clone by 5′-RACE (rapid amplification of cDNA ends) (GenBank accession number DQ507850). The open reading frame (ORF) of GRSF cDNA predicts a protein comprising 207 amino acids with a molecular mass of about 24,000 daltons.

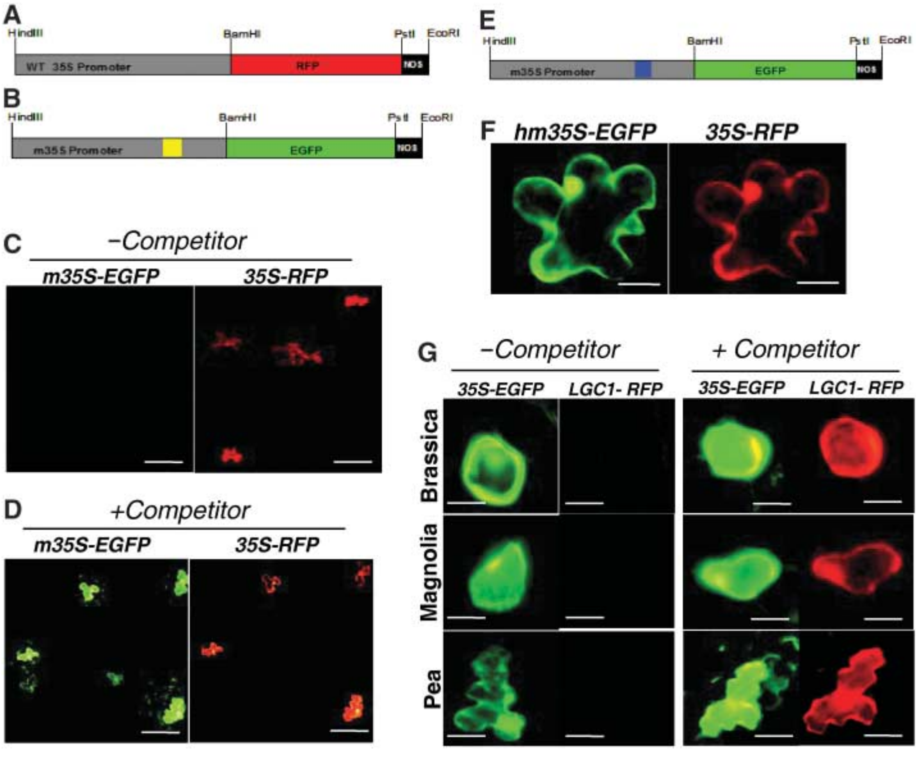


Fig. 2. *CaMV35S* promoter becomes inactive when modified by insertion of silencer region from *LGC1* promoter. This silencing can be reversed by cotransformation with competitor. (**A**) Schematic representation of the wild-type (WT) *CaMV35S-RFP*. Digestion enzymes are noted. (**B**) Schematic representation of the modified *CaMV35S-EGFP* harboring the *LGC1* silencer region. Yellow box: AGATTTATCAGTGGCTGAATTTGGGTGCTGTAGAGACAGAATT. (**C**) Cotransformation of lily petal cells with modified *CaMV35S-EGFP* and wild-type *CaMV35S-RFP* shows expression of RFP only. Scale bar, 200 μm. (**D**) However, cotransformation of modified *CaMV35S-EGFP* with excess competitor (100:1) resulted in its reactivation, resulting in expression of GFP. The same cells show RFP expression due to activity of wild-type *CaMV35S-RFP*. Scale bar, 200 μm. (**E**) Schematic representation of insertion of random human sequence in *CaMV35S-EGFP*. Blue box: TCTCTTACACAGGCAATGATGACATCATCATGACCTCTAAAGA. (**F**) Insertion of random human sequence in *CaMV35S-EGFP* had no effect on expression of GFP; *CaMV35S-RFP* was used as a control. Scale bar, 25 μm. (**G**) *LGC1-RFP* normally inactive becomes activated by excess of competitor (100:1), showing expression of the RFP in petal tissues of diverse plants tested. *CaMV35S-EGFP*, used as an internal control, shows activity irrespective of the presence or absence of competitor. Scale bar: *Brassica*, 10 μm; *Magnolia* and pea, 25 μm.

A BLAST search of deduced amino acid sequences revealed that the C-terminal portion of GRSF exhibits high similarity to nucleolins such as maize nucleic acid–binding protein (NBP) (*8*), *Arabidopsis* nucleolin (*9*), and the potato single-stranded DNA-binding repressor SEBF (*10*). However, the N-terminal region of GRSF shows arginine/serine-rich motifs that are conserved in SON repressor proteins (Fig. 3A and fig. S4). SON and its isoforms are negative regulatory element–binding proteins that have so far been identified only in humans and other mammalian systems (*11*). An AT-hook (*12*) domain, 5-hydroxytryptamine 5B receptor (*13*), and histone H5 signatures (*14*) were also detected in the GRSF sequence (Fig. 3A). Proteins containing AT-hook domains bind minor grooves of A/T-rich sequences and are considered to coregulate transcription by modifying the architecture of DNA by recruiting proteins involved in chromatin remodeling and condensation, thus modifying the architecture of the bound DNA (*12*). In addition, GRSF contains RNA recognition motifs that might mediate nuclear RNA processing activity in addition to its probable transcriptional regulatory role. GRSF contains a domain with the potential to adopt a coiled-coil structure, which has been reported in several transcription factors (*15*). Comparison of the biochemical properties of GRSF to those of nucleolins, ribonucleoproteins, and known plant repressors shows that GRSF has the lowest molecular weight but the highest arginine content (16.4%) and a calculated isoelectric point of 10.40. All these observations point toward GRSF being a novel eukaryotic DNA-binding repressor protein.

GRSF is localized in nuclei of non–male gamete lineage cells. GRSF transcripts are present at high levels in leaf and petal tissues but at moderate levels in pollen and ovary tissues. No signal was detectable in generative cells (Fig. 3B). The positive signal from pollen RNA and the absence of signal from isolated generative cells show that GRSF transcripts are present in the vegetative cells of pollen. Low signal from total pollen RNA is not unexpected, as GRSF is expressed in one cell of pollen only. Although a lower level of GRSF expression in ovary tissues is intriguing, transcript levels are not always tightly linked to cellular levels of protein products (*16*). The deduced amino acid sequence of GRSF contains a bipartite nuclear targeting sequence and a putative arginine-rich nuclear localization signal. Immunolocalization experiments using antibodies to GRSF showed signal in the nucleus of lily uninucleate microspores, in the vegetative cell nucleus of the bicellular stage of pollen development, and in anther wall cells (Fig. 3C); however, no signal was detectable in the generative cell nucleus.

To determine whether GRSF protein targeted chromatin *LGC1* silencer elements in vivo, we performed chromatin immunoprecipitation (ChIP) with antibodies to GRSF. The chromatin fragments that coimmunoprecipitate with GRSF were analyzed by real-time quantitative polymerase chain reaction (PCR) using primers specific for promoter and ORF sequences of *LGC1*, generative cell–specific histones, and the pollen vegetative cell–specific pectate lyase gene. Our results confirmed that GRSF occupies a specific domain in the promoter region of *LGC1* (Fig. 3D). We also analyzed the presence of GRSF on the promoter of the generative cell–specific histone gene *gcH3* (*17*). We observed that *gcH3* also shows specific immunoprecipitation with antibodies to GRSF (Fig. 3D). In addition, our analysis of the pollen-expressed pectate lyase gene as a control showed no association of GRSF with this vegetative cell–expressed gene.

These ChIP results provide direct evidence that promoters of *LGC1* and other generative cell–specific genes are likely targets for GRSF-mediated transcriptional repression. Our results thus show a direct correlation between the recruitment of GRSF to the upstream sequences of specific genes and their male gametic cell–specific expression.

Core silencer sequences are conserved in various male germline–specific genes. We used radiolabeled 43-bp double-stranded silencer sequence oligonucleotide in an EMSA test of its binding to recombinant GRSF (fig. S5). Further EMSA experiments using mutated oligonucleotides containing blocks of 10-bp mutations showed that mutant 3 corresponded to nucleotide sequence critical for the binding of GRSF (Fig. 4A). A series of oligonucleotides that carried blocks of 4-bp mutations were then used as cold competitors with radiolabeled wild-type 43-bp oligonucleotide. All of the mutants except mutants 7 and 8 nearly abolished the binding of labeled wild-type oligonucleotide (Fig. 4B). The partial inhibition of binding by mutants 7 and 8 suggests that the sequences (GGCT and GAAT) altered in these mutants form a component of the optimal binding site for GRSF. The 8-bp sequence motif represented by both mutants 7 and 8 is also contained within the 11-bp repressor-binding motif defined by mutant 3 (Fig. 4A), thus identifying it as the core silencer domain recognized by GRSF.

Our search for similar cis-acting silencers in other male germline–specific genes showed a similar conserved motif with four invariant bases (Fig. 4C). These genes include male gamete–specific histone *gcH3* of lily (*4*,*17*), male gamete–specific histone H3 variant of *Arabidopsis* (*18*), and three additional *Arabidopsis* genes that include *DUO1* (*5*) and *At5g49150* (*6*). It is noteworthy that out of 15 histone H3 genes in the *Arabidopsis* genome (*18*), only the male germline–specific H3 contains the core GRSF-binding domain. The recruitment of GRSF to the silencer motif of lily *LGC1* and *gcH3*, as shown by ChIP assay and the presence of a similar silencer motif in three *Arabidopsis* genes, suggests that they could be direct target genes of GRSF or a similar functionally conserved repressor. Our database search for a GRSF-type repressor indeed showed the presence of similar expressed sequence tags in *Medicago*, maize, rice, *Arabidopsis*, wheat, and *Hordeum* (fig. S6).

Conclusions. Conservation of the repressor-binding site and its associated repressor in phylogenetically distant plants suggests that specific repressor binding element–mediated silencing may be a general mechanism for regulating the expression of male germline–specific genes. Our data show that flowering plant male germline–specific genes are maintained in a repressed state in non–male germline cells via negative transcriptional regulation mediated by GRSF or its functional orthologs that are ubiquitously present in nonmale gametic cells. The presence of GRSF in uninucleate microspores but its absence in one of the daughter cells (the generative cell), with corresponding activation of the male germline–specific transcriptional program, suggests that release from GRSF-imposed repression is a determining event in sperm cell development of flowering plants. Through its regulation of germline-specific genes such as *DUO1* that are essential for gamete development, GRSF may function as a key element of a network of regulatory controls of male gamete development. The presence of a GRSF with conserved binding in the basal angiosperm *Magnolia* suggests that the recruitment of GRSF as a regulatory factor controlling the timing and location of expression of male germline genes might be one of the key processes in the evolution of the reproductive system of flowering plants.

The importance of GRSF in controlling a key developmental event in plant biology is comparable to that of neuron-restrictive silencing factor [NRSF; also known as REST (repressor element–1 silencing transcription factor)] for animal systems. NRSF/REST is an evolutionarily conserved repressor with homologs in various species (*Caenorhabditis elegans*, *Drosophila*, *Xenopus*, mouse, and human)

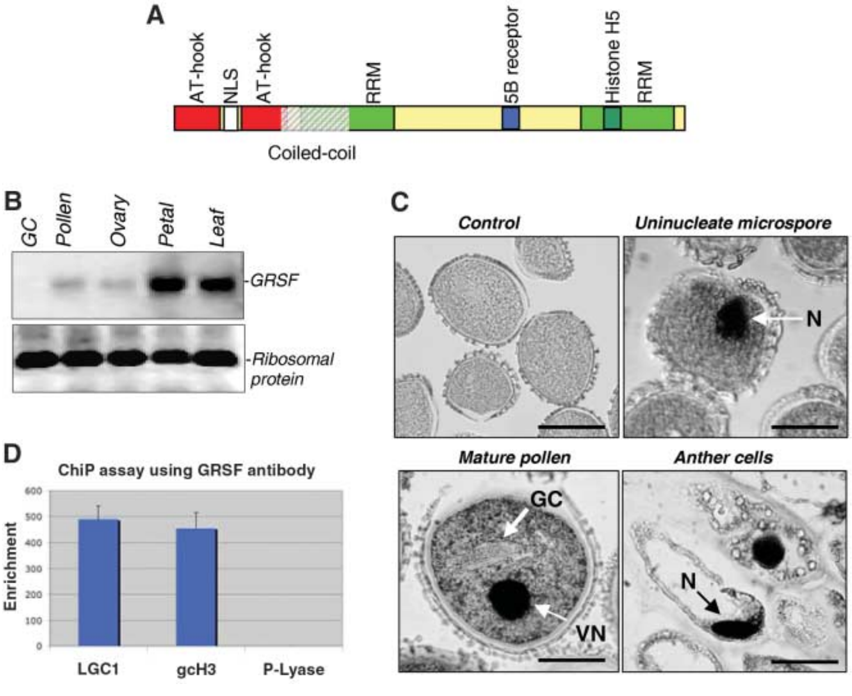


Fig. 3. Identification, cloning, expression, and ChIP analysis of repressor protein with specificity toward the *LGC1* silencer domain. (**A**) Schematic representation of predicted functional domains on the repressor (GRSF). NLS, nuclear localization signal; RRM, RNA binding motifs. The GRSF domain with the potential to adopt a coiled-coil structure is shaded. (**B**) Reverse transcription PCR analysis showing *GRSF* mRNA expression in various lily tissues with the exception of generative cells (GC). (**C**) Nuclear localization of GRSF in the nuclei of uninucleate microspores (N) and the vegetative cell nucleus (VN) of mature bicellular pollen. Anther wall cells also exhibit nuclear localization (N) of GRSF. Scale bars, 100 μm. (**D**) GRSF binds to the silencer region of *LGC1* promoter in vivo. ChIP assay of *LGC1* and generative cell–specific histone *H3* promoter (gcH3) used antibodies to GRSF peptide. The data represent the ratio of the amount of DNA immunoprecipitated using specific antibody to that when antibody to GRSF was omitted, as determined by quantitative real-time PCR. P-lyase, pectate lyase.

that represses transcription from promoters of numerous neuron-specific genes in neural precursors and non-neuronal cells (*19*) (fig. S7). Silencing of neural-specific genes is mediated via recruitment of the corepressor CoREST, which functions as a molecular beacon for the recruitment of specialized silencing machinery (*19*). The question of whether GRSF-induced silencing of male germline–specific genes in the rest of the plant cells involves associated corepressor(s) and the nature of the silencing machinery required for long-term repression remain exciting areas for further investigation.

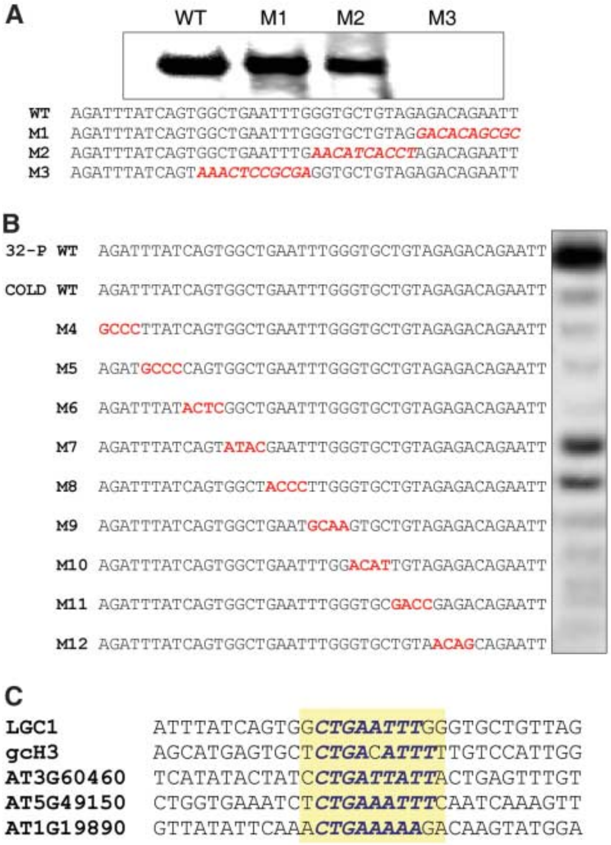

A
WT
M1
M2
M3
WT AGATTTATCAGTGGCTGAATTTGGGTGCTGTAGAGACAGAATT
M1 AGATTTATCAGTGGCTGAATTTGGGTGCTGTAGGACACAGCGC
M2 AGATTTATCAGTGGCTGAATTTGAACATCACCTAGACAGAATT
M3 AGATTTATCAGTAAACTCCGCGAGGTGCTGTAGAGACAGAATT
B
32-P WT AGATTTATCAGTGGCTGAATTTGGGTGCTGTAGAGACAGAATT
COLD WT AGATTTATCAGTGGCTGAATTTGGGTGCTGTAGAGACAGAATT
M4 GCCCTTATCAGTGGCTGAATTTGGGTGCTGTAGAGACAGAATT
M5 AGATGCCCCAGTGGCTGAATTTGGGTGCTGTAGAGACAGAATT
M6 AGATTTATACTCGGCTGAATTTGGGTGCTGTAGAGACAGAATT
M7 AGATTTATCAGTATACGAATTTGGGTGCTGTAGAGACAGAATT
M8 AGATTTATCAGTGGCTACCCTTGGGTGCTGTAGAGACAGAATT
M9 AGATTTATCAGTGGCTGAATGCAAGTGCTGTAGAGACAGAATT
M10 AGATTTATCAGTGGCTGAATTTGGACATTGTAGAGACAGAATT
M11 AGATTTATCAGTGGCTGAATTTGGGTGCGACCGAGACAGAATT
M12 AGATTTATCAGTGGCTGAATTTGGGTGCTGTAACAGCAGAATT
C
LGC1 ATTTATCAGTGGCTGAATTTGGGTGCTGTTAG
gcH3 AGCATGAGTGCTCTGACATTTTTGTCCATTGG
AT3G60460 TCATATACTATCCTGATTATTACTGAGTTTGT
AT5G49150 CTGGTGAAATCTCTGAAATTTCAATCAAAGTT
AT1G19890 GTTATATTCAAACTGAAAAAGACAAGTATGGA

Fig. 4. Identification of core binding domain of GRSF within the silencer region of *LGC1* promoter and conservation of core silencer domain in male germline genes. (**A**) EMSA using recombinant GRSF shows specific binding to 43-bp oligonucleotide sequence of the *LGC1* promoter (WT). Mutations in the region GGCTGAATTT of the oligonucleotide abolished specific binding (M3); mutations in other regions (M1 and M2) had no effect on binding. Mutated sequences are in red. (**B**) LGC1 oligonucleotides (43 bp) carrying 4-bp mutation blocks (marked in red) used as cold competitors in EMSAs with concentration ratios of 100:1. Mutated oligonucleotides 7 and 8 exhibited the lowest capacity to compete with labeled WT probe. An 8-bp sequence covered by these two oligonucleotides lies within the 10-bp region GGCTGAATTT identified by 10-bp block mutations. (**C**) Conservation of GRSF minimal binding site in the promoter regions of lily and *Arabidopsis* male germline–specific genes. *AT1G19890* encodes *Arabidopsis* male germline–specific H3 histone, *AT5G49150* encodes *Arabidopsis* male germline–specific unknown gene, and *AT3G60460* encodes *Arabidopsis DUO1* gene expressed in male germline cells. Core binding domain is shaded in yellow, with conserved sequences marked in blue italics.

References and Notes

1. J. P. Mascarenhas, *Plant Cell* **1**, 657 (1989).
2. T. Mori, H. Kuroiwa, T. Higashiyama, T. Kuroiwa, *Nat. Cell Biol.* **8**, 64 (2006).
3. H. Xu, I. Swoboda, P. L. Bhalla, M. B. Singh, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2554 (1999).
4. H. Xu, I. Swoboda, P. L. Bhalla, M. B. Singh, *Plant Mol. Biol.* **39**, 607 (1999).
5. N. Rotman *et al.*, *Curr. Biol.* **15**, 244 (2005).
6. M. L. Engel, R. Holmes-Davis, S. McCormick, *Plant Physiol.* **138**, 2124 (2005).
7. M. Singh, P. L. Bhalla, H. Xu, M. B. Singh, *FEBS Lett.* **542**, 47 (2003).
8. W. Cook, J. Walker, *Nucleic Acids Res.* **20**, 359 (1992).
9. Y. V. Shidlovskii *et al.*, *EMBO J.* **24**, 97 (2005).
10. B. Boyle, N. Brisson, *Plant Cell* **13**, 2525 (2001).
11. C. Sun *et al.*, *J. Biol. Chem.* **276**, 24059 (2001).
12. A. Siddiqa *et al.*, *Nature* **410**, 383 (2001).
13. C. D. Strader, T. M. Fong, M. R. Tota, D. Underwood, R. A. F. Dixon, *Annu. Rev. Biochem.* **63**, 101 (1994).
14. M. Buttinelli, G. Panetta, D. Rhodes, A. Travers, *Genetica* **106**, 117 (1999).
15. P. Burkhard *et al.*, *Trends Cell Biol.* **11**, 82 (2001).
16. N. Imin, T. Kerim, J. J. Weinmann, B. G. Rolfe, *Mol. Cell. Proteomics* **5**, 274 (2006).
17. T. Okada, P. L. Bhalla, M. B. Singh, *Plant Cell Physiol.* **46**, 797 (2005).
18. T. Okada, M. Endo, M. B. Singh, P. L. Bhalla, *Plant J.* **44**, 557 (2005).
19. V. V. Lunyak, M. G. Rosenfeld, *Cell* **121**, 499 (2005).
20. We thank S. Russell and B. Rolfe for critical reading of the manuscript and suggestions. Supported by the Australian Research Council.

Supporting Online Material
www.sciencemag.org/cgi/content/full/313/5786/496/DC1
Materials and Methods
Figs. S1 to S7
References

30 January 2006; accepted 30 May 2006
10.1126/science.1125526

Violation of Kirchhoff's Laws for a Coherent *RC* Circuit

J. Gabelli,[1] G. Fève,[1] J.-M. Berroir,[1] B. Plaçais,[1] A. Cavanna,[2]
B. Etienne,[2] Y. Jin,[2] D. C. Glattli[1,3]*

What is the complex impedance of a fully coherent quantum resistance-capacitance (*RC*) circuit at gigahertz frequencies in which a resistor and a capacitor are connected in series? While Kirchhoff's laws predict addition of capacitor and resistor impedances, we report on observation of a different behavior. The resistance, here associated with charge relaxation, differs from the usual transport resistance given by the Landauer formula. In particular, for a single-mode conductor, the charge-relaxation resistance is half the resistance quantum, regardless of the transmission of the mode. The new mesoscopic effect reported here is relevant for the dynamical regime of all quantum devices.

For a classical circuit, Kirchhoff's laws prescribe the addition of resistances in series. Its failure has been a central issue in developing our understanding of electronic transport in mesoscopic conductors. Indeed, coherent multiple electronic reflections between scatterers in the conductor were found to make the conductance nonlocal (*1*, *2*). A new composition law of individual scatterer contribution to resistance was found that led to the solution of the problem of electron localization (*3*) and, later, to formulation of the electronic conduction in terms of scattering of electronic waves (*4*). Nonadditivity of series resistances, or of parallel conductances, nonlocal effects, and negative four-point resistances (*5*) have been observed in a series of transport experiments at low temperature, where phase coherence extends over the mesoscopic scale (*6*, *7*). It is generally accepted that the conductance of a phase-coherent quantum conductor is given by the Landauer formula and its generalization to multilead conductors (*8*), which relate the conductance to the transmission of electronic waves by the conductance quantum e^2/h. But, how far is this description robust at finite frequency, where conductance combines with nondissipative circuit elements such as capacitors or inductors? Are there significant departures from the dc result? The question is important, as recent advances in quantum information highlight the need for fast manipulation of quantum systems, in particular quantum conductors. High-frequency quantum transport has been theoretically addressed, showing that a quantum *RC* circuit displays discrepancies with its classical counterpart (*9*, *10*). It was shown that a counterintuitive modification of the series resistance led to the situation in which the resistance is no longer described by the Landauer formula and does not depend on transmission in a direct way (*9*, *10*). Instead, it is directly related to the dwell time of electrons in the capacitor. Moreover, when the resistor transmits in a single electronic mode, a constant resistance was found, equal to the half-resistance quantum $h/2e^2$, i.e., it was not transmission-dependent. This resistance, modified by the presence of the coherent capacitor, was termed a "charge-relaxation resistance" to distinguish it from the usual dc resistance, which is sandwiched between macroscopic reservoirs and described by the Landauer formula. The quantum charge–relaxation resistance, as well as its generalization in nonequilibrium systems, is an important concept that can be applied to quantum information. For example, it enters into the problem of quantum-limited detection of charge qubits (*11*, *12*), in the study of high-frequency-charge quantum noise (*13*–*15*), or in the study of dephasing of an electronic quantum interferometer (*16*). In molecular electronics, the charge-relaxation resistance is also relevant to the THz frequency response of systems such as carbon nanotubes (*17*).

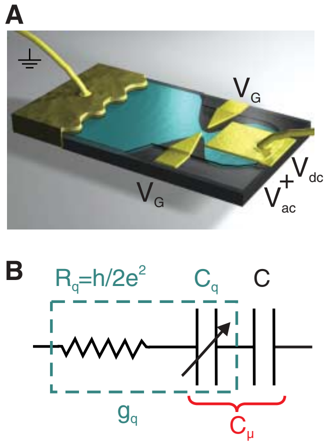


Fig. 1. The quantum capacitor realized using a 2DEG (**A**) and its equivalent circuit (**B**). The capacitor consists of a metallic electrode (in gold) on top of a submicrometer 2DEG quantum dot (in blue) defining the second electrode. The resistor is a QPC linking the dot to a wide 2DEG reservoir (in blue), itself connected to a metallic contact (dark gold). The QPC voltage V_G controls the number of electronic modes and their transmission. The radio frequency voltage V_{ac}, and eventually a dc voltage V_{dc}, are applied to the counter-electrode, whereas the ac current, from which the complex conductance is deduced, is collected at the ohmic contact. As predicted by the theory, the relaxation resistance R_q, which enters the equivalent circuit for the coherent conductance, is transmission-independent and equal to half the resistance quantum. The capacitance is the serial combination C_μ of the quantum and the geometrical capacitances (C_q and C, respectively). C_q is transmission-dependent and strongly modulated by V_{dc} and/or V_G. The combination of R_q and C_q forms the impedance $1/g_q$ of the coherent quantum conductor.

We report on the observation and quantitative measurement of the quantum charge–relaxation resistance in a coherent *RC* circuit realized in a two-dimensional electron gas (2DEG) (see Fig. 1A). The capacitor is made of a macroscopic metallic electrode on top of a 2DEG submicrometer dot defining the second electrode. The resistor is a quantum point contact (QPC) connecting the dot to a wide 2DEG macroscopic reservoir. We address the coherent regime in which electrons emitted from the reservoir to the dot are backscattered without loss of coherence. In this regime, we have checked the prediction made in refs. (*9*, *10*) that the charge-relaxation resistance is not given by the Landauer formula resistance but instead is constant and equals $h/2e^2$, as the QPC transmission is varied. Note that we consider here a spin-polarized regime and that the factor 1/2 is not the effect of spin, but a hallmark of a charge-relaxation resistance. When coherence is washed out by thermal broadening, the more conventional regime pertaining to dc transport is recovered. The present work differs from previous capacitance measurements where, for spectroscopic purpose, the dot reservoir coupling was weak and the ac transport regime was incoherent (*18*, *19*). As a consequence, although quantum effects in the capacitance were observable, the quantum charge–relaxation resistance was not accessible in these earlier experiments.

At zero temperature in the coherent regime and when a single mode is transmitted by the QPC, the mesoscopic *RC* circuit is represented by the equivalent circuit of Fig. 1B (*9*, *10*). The geometrical capacitance *C* is in series with the quantum admittance $g_q(\omega)$ connecting the ac current flowing in the QPC to the ac internal potential of the dot:

$$g_q(\omega) = 1/\left(\frac{h}{2e^2} + \frac{1}{-i\omega C_q}\right), (T = 0) \quad (1)$$

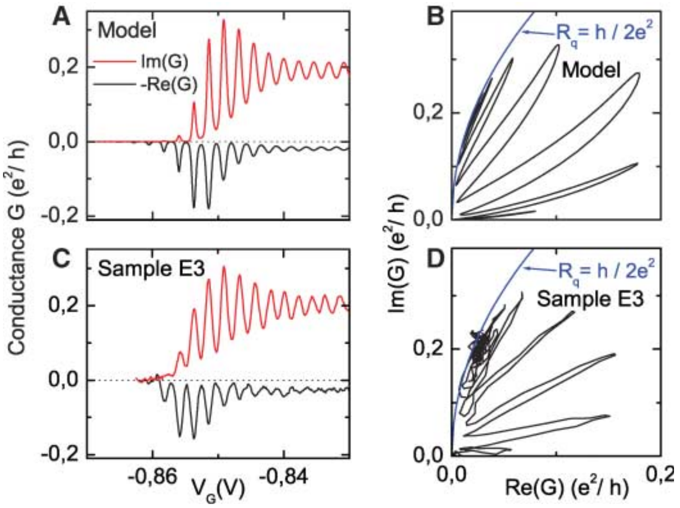
The nonlocal quantum impedance behaves as if it were the series addition of a quantum capacitance C_q with a constant contact resistance $h/2e^2$. $C_q = e^2(dN/d\varepsilon)$ is associated with the local density of state $dN/d\varepsilon$ of the mode propagating in the dot, taken at the Fermi energy. The striking effect of phase coherence is that the QPC transmission probability *D* affects the quantum capacitance (see Eq. 4) but not the resistance. The total circuit admittance *G* is simply:

$$G = \frac{-i\omega C g_q(\omega)}{-i\omega C + g_q(\omega)} = \frac{-i\omega C_\mu(2e^2/h)}{-i\omega C_\mu + (2e^2/h)}, \quad (T = 0) \quad (2)$$

where $C_\mu = CC_q/(C + C_q)$ is the electrochemical capacitance. In the incoherent regime, both resistance and quantum capacitance vary with transmission. The dot forms a second reservoir and the electrochemical capacitance C_μ is in series with the QPC resistance *R*. In particular, when the temperature is high enough to smooth the capacitor density of states, the Landauer formula $R = h/e^2 \times 1/D$ is recovered.

Several samples have been measured at low temperatures, down to 30 mK, which show analogous features. We present results on two samples made with 2DEG defined in the same high-mobility GaAsAl/GaAs heterojunction, with nominal density $n_s = 1.7 \times 10^{15}$ m^{-2} and mobility $\mu = 260$ V^{-1} m^2 s^{-1}. A finite magnetic field ($B = 1.3$ T)

Fig. 2. Complex conductance of sample E3 as function of the gate voltage V_G for $T = 100$ mK and $\omega/2\pi = 1.2$ GHz, at the opening of the first conduction channel (**C**) and its Nyquist representation in (**D**). The theoretical circle characteristic of the coherent regime is shown as a solid line. (**A** and **B**) show the corresponding curves for the simulation of sample E3 using the 1D model with $C = 4fF$, $C_\mu = 1fF$.



[1] Laboratoire Pierre Aigrain, Département de Physique de l'Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris Cedex 05, France. [2] Laboratoire de Photonique et Nanostructures, UPR20 CNRS, Route de Nozay, 91460 Marcoussis Cedex, France. [3] Service de Physique de l'Etat Condensé (CNRS URA 2464), Commissariat à l'Energie Atomique (CEA) Saclay, F91191 Gif-sur-Yvette, France.

*To whom correspondence should be addressed. E-mail: glattli@lpa.ens.fr

is applied, so as to work in the ballistic quantum Hall regime with no spin degeneracy (*20*).

The real and imaginary parts of the admittance *Im*(*G*) and *Re*(*G*) as a function of QPC gate voltage V_G at the opening of the first conduction channel are shown in Fig. 2C. On increasing V_G, we can distinguish three regimes. At very negative $V_G \leq -0.86$ V, the admittance is zero. Starting from this pinched state, peaks are observed in both *Im*(*G*) and *Re*(*G*). Following a maximum in the oscillations, a third regime occurs where *Im*(*G*) oscillates nearly symmetrically about a plateau, whereas the oscillation amplitude decreases smoothly. Simultaneously, peaks in *Re*(*G*) quickly disappear to vanish in the noise.

Comparing these observations with the results of refs. (*9*,*10*), using a simplified one-dimensional (1D) model for C_q with one conduction mode and a constant energy level spacing in the dot Δ (*21*), the simulation (Fig. 2A) shows a striking similarity to the experimental conductance traces in Fig. 2C. In this simulation, V_G determines the transmission *D* but also controls linearly the 1D dot potential. The transmission is chosen to vary with V_G according to a Fermi-Dirac–like dependence appropriate to describe QPC transmission (*22*). This model can be used to get a better understanding of the different conductance regimes. Denoting *r* and *t* the amplitude reflection and transmission coefficients of the QPC ($r^2 = 1 - D$, $t = \sqrt{D}$), we first calculate the scattering amplitude of the *RC* circuit:

$$s(\varepsilon) = r - t^2 e^{i\varphi}\sum_{n=0}^{\infty}(re^{i\varphi})^n = \frac{r - e^{i\varphi}}{1 - re^{i\varphi}} \qquad (3)$$

where ε is the Fermi energy relative to the dot potential and φ =2πε/Δ is the phase accumulated for a single turn in the quantum dot. The zero-temperature quantum capacitance is then given by:

$$Cq = e^2(dN/d\varepsilon) = \frac{1}{2i\pi}s^+\frac{\partial s}{\partial \varepsilon} = \frac{e^2}{\Delta} \times \frac{1 - r^2}{1 - 2r\cos(2\pi\varepsilon/\Delta) + r^2} \qquad (4)$$

Therefore, C_q exhibits oscillations when the dot potential is varied. When $r \rightarrow 0$, these oscillations vanish and $C_q \rightarrow e^2/\Delta$. As reflection increases, oscillations are growing with maxima $(e^2/\Delta)[(1 + r)/(1 - r)]$ and minima $(e^2/\Delta)[(1 - r)/(1 + r)]$. For strong reflection, Eq. 4 gives resonant Lorentzian peaks with an energy width $D\Delta/2$ given by the escape rate of the dot. However, at finite temperature, the conductance in Eq. 1 has to be thermally averaged to take into account the finite energy width of the electron source so that:

$$g_q(\omega) = \int d\varepsilon\left(-\frac{\partial f}{\partial \varepsilon}\right) \times \frac{1}{h/2e^2 + 1/(-i\omega C_q)}, \ (T \neq 0) \qquad (5)$$

where *f* is the Fermi-Dirac distribution. Again the nonlocal admittance behaves as if it were the serial association of a charge-relaxation resistance R_q and a capacitance that we still denote C_q. In the weak transmission regime ($D \rightarrow 0$), when $D\Delta \ll k_BT$, Eq. 5 yields thermally broadened capacitance peaks with

$$C_q \approx \frac{e^2}{4k_BT\cosh^2(\delta\varepsilon/2k_BT)}, \ (D \ll 1) \qquad (6)$$

where δε denotes the energy distance to a resonant dot level. Note that these capacitance peaks do not depend on the dot parameters and can be used as a primary thermometer. Similar but transmission-dependent peaks are predicted in the inverse resistance

$$1/R_q \approx D\frac{e^2}{h} \times \frac{\Delta}{4k_BT\cosh^2(\delta\varepsilon/2k_BT)}, (D \ll 1) \qquad (7)$$

This result is reminiscent of the thermally broadened resonant tunneling conductance for transport through a quantum dot. A consequence of the finite temperature is the fact that the resistance is no longer constant. This thermally induced divergence of R_q at low transmission restores a frequency-dependent pinch-off for $R_q \gg 1/C_q\omega$, as can be seen in both model and experiment in Fig. 2, A and C. As mentioned above, for $k_BT \gg \Delta$, the quantum dot looks like a reservoir and the Landauer formula is recovered.

The coherent and the thermally broadened regimes are best demonstrated in the Nyquist representation *Im*(*G*) versus *Re*(*G*) of the experimental data in Fig. 2D. This representation allows us to easily distinguish constant resistance from constant capacitance regimes, as they correspond to circles respectively centered on the real and imaginary axis. Whereas, for low transmission, the Nyquist diagram strongly depends on transmission, the conductance oscillations observed in Fig. 2C collapse on a single curve in the coherent regime. Moreover, this curve is the constant $R_q = h/2e^2$ circle. By contrast, admittance peaks at low transmission correspond to a series of lobes in the Nyquist diagram, with slopes increasing with transmission in qualitative agreement with Eqs. 6 and 7. These lobes and the constant R_q regime are well reproduced by the simulations in Fig. 2B. Here, the value of C_μ and the electronic temperature are deduced from measurement. In our experimental

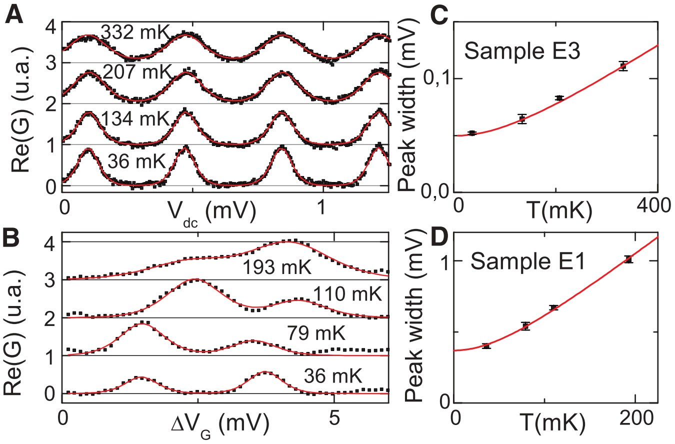


Fig. 3. Coulomb-blockade oscillations in the real part of the ac conductance in the low-transmission regime. The control voltage is applied to the counter-electrode for sample E3 (**A**) and to the QPC gate for sample E1 (**B**). The temperature dependence is used for absolute calibration of our setup, as described in the text: The peak width, shown in (**C** and **D**) as a function of temperature, is deduced from theoretical fits (solid lines) using Eq. 7 and taking a linear dependence of energy with the control voltage. Lines in (C) and (D) are fits of the experimental results using a $\sqrt{(T^2 + T_0^2)}$ law to take into account a finite residual electronic temperature T_0.

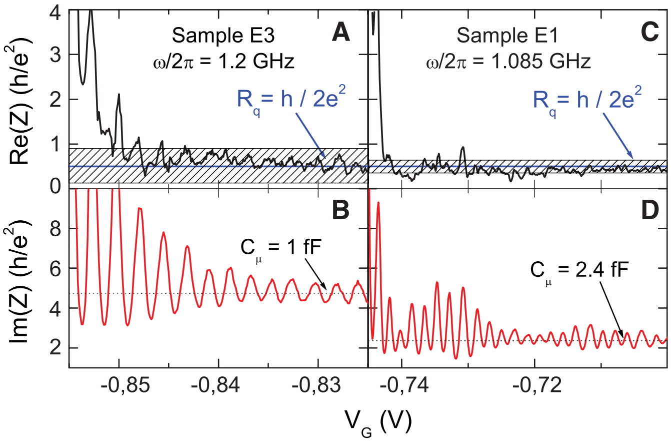


Fig. 4. Complex impedance of sample E3 (**A** and **B**) and sample E1 (**C** and **D**) as a function of QPC voltage for *T* = 30 mK and *B* =1.3 T. The dashed lines in (B and D) correspond to the values of C_μ deduced from calibration. The horizontal solid lines in (A and C) indicate the half-quantum of resistance expected for the coherent regime. Uncertainties on R_q are displayed as hatched areas.

conditions, the simulated traces are virtually free of adjustable parameters as $C \geq 4C_\mu \gg C_q$.

It is important to note that in a real system, the weak transmission regime is accompanied by Coulomb blockade effects that are not taken into account in the above model. In the weak transmission regime and $T = 0$, using an elastic cotunneling approach (*23*, *24*), we have checked that there is no qualitative change except for the energy scale that now includes the charging energy so that Δ is replaced by $\Delta + e^2/C = e^2/C_\mu$. At large transmission, the problem is nonperturbative in tunnel coupling and highly nontrivial. Calculations of the thermodynamic capacitance exist [(*25*, *26*), and (*27*) plus references therein], but at present, no comprehensive model is available that would include both charge-relaxation resistance and quantum capacitance for finite temperature and/or large transmission.

Calibration of our admittance measurements is a crucial step toward extracting the absolute value of the constant charge-relaxation resistance. As at GHz frequencies, direct calibration of the whole detection chain is hardly better than 3 dB, we shall use here an indirect, but absolute, method, often used in Coulomb blockade spectroscopy, that relies on the comparison between the gate voltage width of a thermally broadened Coulomb peak ($\propto k_BT$) and the Coulomb peak spacing ($\propto e^2/C_\mu$). From this, an absolute value of C_μ can be obtained. The real part of the admittance of sample E3 is shown as a function of the dc voltage V_{dc} at the counter-electrode, for a given low transmission (Fig. 3A). A series of peaks with periodicity $\Delta V_{dc} = 370\ \mu V$ are observed, with the peaks accurately fitted by Eq. 7. Their width, proportional to the electronic temperature T_{el}, is plotted versus the refrigerator temperature T (see Fig. 3C). When corrected for apparent electron heating arising from gaussian environmental charge noise, and if we assume $T_{el} = \sqrt{(T^2 + T_0^2)}$, the energy calibration of the gate voltage yields C_μ and the amplitude $1/C_\mu\omega$ of the conductance plateau in Fig. 2. A similar analysis is done in Fig. 3, B and D, for sample E1 using V_G to control the dot potential. Here, peaks are distorted because of a transmission-dependent background and show a larger periodicity $\Delta V_G = 2$ mV, which reflects the weaker electrostatic coupling to the 2DEG.

Finally, after numerical inversion of the conductance data, we can separate the complex impedance into the contributions of the capacitance, $1/C_\mu\omega$, and the relaxation resistance R_q. The results in Fig. 4 demonstrate deviations from standard Kirschhoff's laws: The charge-relaxation resistance R_q remains constant in the regime where the quantum capacitance exhibits strong transmission-dependent oscillations; this constant value equals, within experimental uncertainty, half the resistance quantum as prescribed by theory (*9*, *10*). In the weak transmission regime, the Landauer formula is recovered because of thermal broadening, and R_q diverges as it does in the dc regime. Furthermore, additional measurements at 4 K prove that the classical behavior is indeed recovered in the whole transmission range whenever $k_BT \gg e^2/C_\mu$.

In conclusion, we have experimentally shown that the series association of a quantum capacitor and a model quantum resistor leads to a violation of the dynamical Kirchhoff's law of impedance addition. In the fully coherent regime, the quantum resistor is no longer given by the Landauer formula but by the half-quantized charge-relaxation resistance predicted in refs. (*9*, *10*).

References and Notes

1. R. Landauer, *IBM J. Res. Dev.* **1**, 233 (1957).
2. R. Landauer, *Phil. Mag.* **21**, 863 (1970).
3. P. W. Anderson, *Phys. Rev. B* **23**, 4828 (1981).
4. M. Büttiker, Y. Imry, R. Landauer, S. Pinhas, *Phys. Rev. B* **31**, 6207 (1985).
5. B. Gao, A. Komnik, R. Egger, D. C. Glattli, A. Bachtold, *Phys. Rev. Lett.* **92**, 216804 (2004).
6. See for a review, S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge Univ. Press, Cambridge, 1997).
7. See for a review, Y. Imry, *Introduction to Mesoscopic Physics* (Oxford Univ. Press, Oxford, 1997).
8. M. Büttiker, *Phys. Rev. Lett.* **57**, 1761 (1986).
9. M. Büttiker, A. Prêtre, H. Thomas, *Phys. Rev. Lett.* **70**, 4114 (1993).
10. A. Prêtre, H. Thomas, M. Büttiker, *Phys. Rev. B* **54**, 8130 (1996).
11. S. Pilgram, M. Büttiker, *Phys. Rev. Lett.* **89**, 200401 (2002).
12. A. A. Clerk, S. M. Girvin, A. D. Stone, *Phys. Rev. B* **67**, 165324 (2003).
13. M. Büttiker, H. Thomas, A. Prêtre, *Phys. Lett. A* **180**, 364 (1993).
14. Ya. M. Blanter, M. Büttiker, *Phys. Rep.* **336**, 1 (2000).
15. F. W. J. Hekking, J. P. Pekola, *Phys. Rev. Lett.* **96**, 056603 (2006).
16. G. Seelig, S. Pilgram, A. N. Jordan, M. Büttiker, *Phys. Rev. B* **68**, 161310 (2003).
17. P. J. Burke, *IEEE T. Nanotechnol.* **2** (1), 55 (2003).
18. R. C. Ashoori *et al.*, *Phys. Rev. Lett.* **68**, 3088 (1992).
19. R. C. Ashoori *et al.*, *Phys. Rev. Lett.* **71**, 613 (1992).
20. Materials and methods are available as supporting material on *Science* Online.
21. J. Gabelli, thesis, Université Pierre et Marie Curie, Paris, 2006; online access at (http://tel.ccsd.cnrs.fr/tel00011619).
22. M. Büttiker, *Phys. Rev. B* **41**, 7906 (1990).
23. D. V. Averin, Y. Nazarov, in *Single Charge Tunneling* (NATO ASI ser. B, vol. 294, Plenum, New York, 1992), chap. 6.
24. D. C. Glattli, *Physica B* **189**, 88 (1993).
25. K. A. Matveev, *Phys. Rev. B* **51**, 1743 (1995).
26. L. I. Glazman, I. L. Aleiner, *Phys. Rev. B* **57**, 9608 (1998).
27. P. W. Brouwer, A. Lamacraft, K. Flensberg, *Phys. Rev. B* **72**, 075316 (2005).

28. The Laboratoire Pierre Aigrain is the CNRS-ENS UMR8551 associated with universities Paris 6 and Paris 7. The research has been supported by AC-Nanoscience, SESAME grants, and ANR-05-NANO-028.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1126940/DC1
Materials and Methods

2 March 2006; accepted 26 June 2006
Published online 13 July 2006;
10.1126/science.1126940
Include this information when citing this paper.

Second-Harmonic Generation from Magnetic Metamaterials

Matthias W. Klein,[1,2] Christian Enkrich,[1,2] Martin Wegener,[1,2,3] Stefan Linden[1,2,3]*

We observe second-harmonic generation from metamaterials composed of split-ring resonators excited at 1.5-micrometer wavelength. Much larger signals are detected when magnetic-dipole resonances are excited, as compared with purely electric-dipole resonances. The experiments are consistent with calculations based on the magnetic component of the Lorentz force exerted on metal electrons—an intrinsic second-harmonic generation mechanism that plays no role in natural materials. This unusual mechanism becomes relevant in our work as a result of the enhancement and the orientation of the local magnetic fields associated with the magnetic-dipole resonances of the split-ring resonators.

The concept of metamaterials has changed the spirit of optics and photonics. Researchers no longer just study the rich variety of materials provided by nature but have rather become creative designers who tailor optical properties at will, leading to qualitatively new and unprecedented behavior (*1*–*11*). The key is the nanofabrication of metallic subwavelength-scale functional building blocks, photonic atoms, which are densely packed into an effective material. To a large extent, this emerging field has been stimulated by the 1999 theoretical work of John Pendry's group (*1*), which made two distinct predictions: (i) They proposed split-ring resonators as photonic atoms that could lead to magnetism at optical frequencies—a prerequisite for negative-index metamaterials. (ii) Furthermore, they predicted that enhanced and novel nonlinear-optical properties could arise from such metamaterials. Although aspect (i) has attracted substantial attention from both experiment (*3*–*7*, *12*, *13*) and theory (*14*–*16*) in recent years, aspect (ii) has not, to the best of our knowledge. Experiments have not been reported, nor has a complete consistent microscopic theory of the nonlinear optics of metamaterials been evaluated. This lack of re-

[1]Institut für Angewandte Physik, Universität Karlsruhe (TH), Wolfgang-Gaede-Straße 1, D-76131 Karlsruhe, Germany. [2]DFG-Center for Functional Nanostructures (CFN), Universität Karlsruhe (TH), D-76128 Karlsruhe, Germany. [3]Institut für Nanotechnologie, Forschungszentrum Karlsruhe in der Helmholtz-Gemeinschaft, D-76021 Karlsruhe, Germany.

*To whom correspondence should be addressed. E-mail: stefan.linden@physik.uni-karlsruhe.de

search is contrasted by the potential of metamaterials for giant nonlinear-optical response through local-field enhancements (*1*, *15*) and/or novel mechanisms. This avenue could, for example, lead to ultracompact frequency-doubling devices. Ultimately, parametric nonlinear-optical processes (*17*) in metamaterials might possibly even be used as a gain mechanism, compensating the metamaterial absorption losses in, for example, "perfect lenses" (*2*, *18*). Here, we make steps in direction (ii) and study the lowest-order nonlinear-optical process, that is, second-harmonic generation (SHG). We directly compare our experimental findings on arrays of gold split-ring resonators (SRRs) with the results of a simple theory based on the magnetic component of the Lorentz force on metal electrons in the SRRs—an intrinsic second-harmonic generation mechanism that plays no role in natural materials.

In fact, in usual nonlinear optics, the magnetic-field component of the electromagnetic light wave hardly plays any immediate role at all. Rather, electric dipoles are excited by the electric-field component of the light only. There are, however, some exotic cases of nonlinear optics governed by the magnetic component that we would like to recall before describing our own work. Generally, the magnetic field $\mathbf{B}$ enters by means of the magnetic component of the Lorentz force, $\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$, where q and $\mathbf{v}$ are the electron charge and velocity, respectively. Although the modulus of the $q(\mathbf{v} \times \mathbf{B})$ term becomes comparable in strength to the electric component of the Lorentz force $q\mathbf{E}$ only for relativistic velocities $\mathbf{v}$, it has measurable consequences for optical frequencies at much smaller velocities. For example, it can lead to the photon-drag effect (*19*), a drift velocity of free crystal electrons that is proportional to the intensity of light and directed along the wave vector of light. This effect is employed in commercially available infrared photon-drag p-type germanium photodetectors and can be interpreted as the dynamic Hall effect (*20*). On the same order of perturbation theory in the incident fields, one also gets an oscillatory electron motion at twice the exciting light frequency (*21*). The polarization of the resulting SHG is again directed along the incident wave vector of light. Such SHG has been observed for free vacuum electrons and is called nonlinear Thomson scattering (*19*, *21*) or Larmor radiation. All these nonlinear contributions point in the direction of the incident wave vector of light $\mathbf{K} \sim (\mathbf{v} \times \mathbf{B}) \sim (\mathbf{E} \times \mathbf{B})$; they are longitudinal components, which cannot propagate in the forward direction. In contrast, a significant component of the local magnetic field pointing along the direction of the incident wave vector of light, as in some of our SRR structures below, could clearly lead to a transverse component of SHG. A transverse component, in contrast to a longitudinal component, can efficiently be radiated into the far-field forward direction. For the symmetry of the structures to be discussed below, this transverse SHG would be polarized orthogonal to the electric-field component of the incident light, hence perpendicular to the incident linear polarization.

The metamaterials under investigation are planar arrays of SRRs. Each SRR can be viewed as a small *LC*-oscillator circuit. The open ends of a nonmagnetic gold wire form the capacitance *C*; the wire itself is a fraction of one winding of a magnetic coil with inductance *L* (see insets in Fig. 1). The corresponding magnetic dipole moment is oriented perpendicular to the plane of the SRR. Details of design, fabrication and linear-optical characterization of the magnetic metamaterials used here have been described elsewhere (*5*, *6*). Briefly, 100-μm by 100-μm two-dimensional arrays of gold split rings with variable lateral size *l* and thickness *t* = 25 nm, on a square lattice with variable lattice constant *a*, are fabricated on glass substrates coated with a 5-nm thin film of indium-tin-oxide (ITO) with standard electron-beam lithography. The eigenfrequency of our *LC* circuit scales inversely with SRR size, provided the eigenfrequency is much smaller than the bulk metal plasma frequency. For normal incidence and horizontal polarization, the electric field of the light can couple to the capacitance (*5*, *6*), inducing a circulating current in the coil, leading to an oscillatory magnetic dipole moment perpendicular to the SRR plane. This resonant circulating current leads to a resonant enhancement of the local magnetic fields. For vertical incident polarization, neither the electric nor the magnetic component of the light can couple to the *LC* resonance. For both linear polarizations one can, however, excite the SRR Mie resonance, which is located at frequencies higher than that of the *LC* resonance. When exciting the Mie resonance with horizontal incident polarization, the current in the SRR bottom arm is accompanied by currents in the two vertical SRR arms. The latter two oscillate 180 degrees out of phase; hence, one gets a nonzero magnetic-dipole moment (*6*). In contrast, for vertical incident polarization, the response of the Mie resonance is purely electric. Corresponding measured transmittance spectra of the samples investigated here are shown in Fig. 1. The various observed transmittance dips correspond to the resonances discussed above. These two samples as well as others are located on the same glass substrate and have been fabricated in one run.

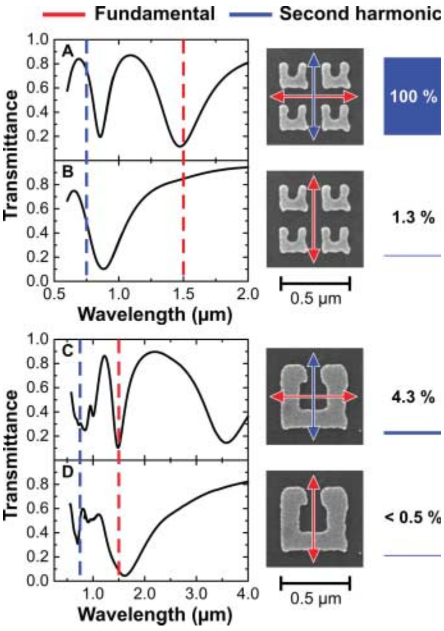


Fig. 1. Summary of measured linear-optical spectra (black solid curves), shown for two relevant magnetic metamaterial samples located on the same substrate. The polarization of the incident light is indicated by the red arrows in the electron micrographs of the corresponding structures. The wavelengths of the exciting light (red) and that of the SHG (blue) are indicated by dashed lines. (**A**) and (**B**) correspond to an array with small SRRs (*l* = 220 nm, *a* = 305 nm), (**C**) and (**D**) to an array with large SRRs (*l* = 480 nm, *a* = 630 nm). The blue bars highlight the corresponding measured SHG signal strengths, normalized to 100% for the strongest SHG signal obtained from the fundamental magnetic (or *LC*) resonance. The detection noise is about 0.2%. The approximate polarization of the SHG emission is indicated by the blue arrows (see also Fig. 2B).

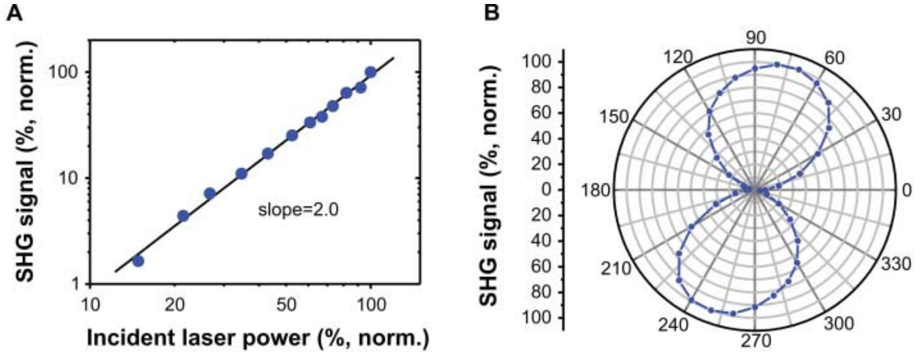


Fig. 2. (**A**) Normalized SHG signal strength versus normalized incident laser power on a log-log scale (for the fundamental magnetic resonance in Fig. 1A). The straight line has a slope of two, as expected for SHG. (**B**) Measured polarization of the SHG emission represented as a polar diagram, oriented as the electron micrograph in Fig. 1A.

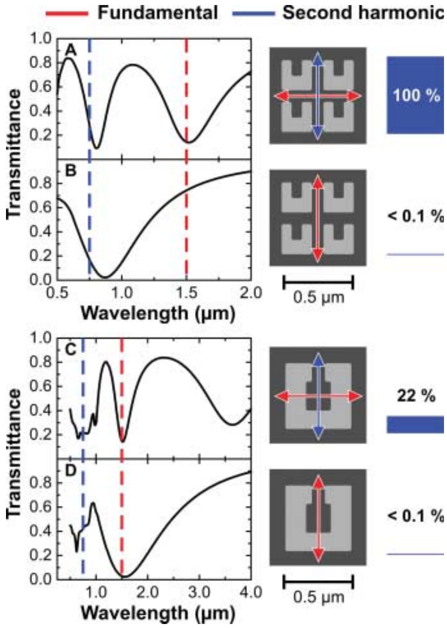


Fig. 3. Theory, presented as the experiment (see Fig. 1). The SHG source is the magnetic component of the Lorentz force on metal electrons in the SRRs.

The setup for measuring the SHG is described in the supporting online material (*22*). We expect that the SHG strongly depends on the resonance that is excited. Obviously, the incident polarization and the detuning of the laser wavelength from the resonance are of particular interest. One possibility for controlling the detuning is to change the laser wavelength for a given sample, which is difficult because of the extremely broad tuning range required. Thus, we follow an alternative route, lithographic tuning (in which the incident laser wavelength of 1.5 μm, as well as the detection system, remains fixed), and tune the resonance positions by changing the SRR size. In this manner, we can also guarantee that the optical properties of the SRR constituent materials are identical for all configurations. The blue bars in Fig. 1 summarize the measured SHG signals. For excitation of the *LC* resonance in Fig. 1A (horizontal incident polarization), we find an SHG signal that is 500 times above the noise level. As expected for SHG, this signal closely scales with the square of the incident power (Fig. 2A). The polarization of the SHG emission is nearly vertical (Fig. 2B). The small angle with respect to the vertical is due to deviations from perfect mirror symmetry of the SRRs (see electron micrographs in Fig. 1). Small detuning of the *LC* resonance toward smaller wavelength (i.e., to 1.3-μm wavelength) reduces the SHG signal strength from 100% to 20%. For excitation of the Mie resonance with vertical incident polarization in Fig. 1D, we find a small signal just above the noise level. For excitation of the Mie resonance with horizontal incident polarization in Fig. 1C, a small but significant SHG emission is found, which is again polarized nearly vertically. For completeness, Fig. 1B shows the off-resonant case for the smaller SRRs for vertical incident polarization.

Although these results are compatible with the known selection rules of surface SHG from usual nonlinear optics (*23*), these selection rules do not explain the mechanism of SHG. Following our above argumentation on the magnetic component of the Lorentz force, we numerically calculate first the linear electric and magnetic field distributions (*22*); from these fields, we compute the electron velocities and the Lorentz-force field (fig. S1). In the spirit of a metamaterial, the transverse component of the Lorentz-force field can be spatially averaged over the volume of the unit cell of size *a* by *a* by *t*. This procedure delivers the driving force for the transverse SHG polarization. As usual, the SHG intensity is proportional to the square modulus of the nonlinear electron displacement. Thus, the SHG strength is expected to be proportional to the square modulus of the driving force, and the SHG polarization is directed along the driving-force vector. Corresponding results are summarized in Fig. 3 in the same arrangement as Fig. 1 to allow for a direct comparison between experiment and theory. The agreement is generally good, both for linear optics and for SHG. In particular, we find a much larger SHG signal for excitation of those two resonances (Fig. 3, A and C), which are related to a finite magnetic-dipole moment (perpendicular to the SRR plane) as compared with the purely electric Mie resonance (Figs. 1D and 3D), despite the fact that its oscillator strength in the linear spectrum is comparable. The SHG polarization in the theory is strictly vertical for all resonances. Quantitative deviations between the SHG signal strengths of experiment and theory, respectively, are probably due to the simplified SRR shape assumed in our calculations and/or due to the simplicity of our modeling. A systematic microscopic theory of the nonlinear optical properties of metallic metamaterials would be highly desirable but is currently not available.

References and Notes

1. J. B. Pendry, A. J. Holden, D. J. Robbins, W. J. Stewart, *IEEE Trans. Microw. Theory Tech.* **47**, 2075 (1999).
2. J. B. Pendry, *Phys. Rev. Lett.* **85**, 3966 (2000).
3. R. A. Shelby, D. R. Smith, S. Schultz, *Science* **292**, 77 (2001).
4. T. J. Yen *et al.*, *Science* **303**, 1494 (2004).
5. S. Linden *et al.*, *Science* **306**, 1351 (2004).
6. C. Enkrich *et al.*, *Phys. Rev. Lett.* **95**, 203901 (2005).
7. A. N. Grigorenko *et al.*, *Nature* **438**, 335 (2005).
8. G. Dolling, M. Wegener, S. Linden, C. Hormann, *Opt. Express* **14**, 1842 (2006).
9. G. Dolling, C. Enkrich, M. Wegener, C. M. Soukoulis, S. Linden, *Science* **312**, 892 (2006).
10. J. B. Pendry, D. Schurig, D. R. Smith, *Science* **312**, 1780; published online 25 May 2006.
11. U. Leonhardt, *Science* **312**, 1777 (2006); published online 25 May 2006.
12. M. W. Klein, C. Enkrich, M. Wegener, C. M. Soukoulis, S. Linden, *Opt. Lett.* **31**, 1259 (2006).
13. W. J. Padilla, A. J. Taylor, C. Highstrete, M. Lee, R. D. Averitt, *Phys. Rev. Lett.* **96**, 107401 (2006).
14. D. R. Smith, S. Schultz, P. Markos, C. M. Soukoulis, *Phys. Rev. B* **65**, 195104 (2002).
15. S. O'Brien, D. McPeake, S. A. Ramakrishna, J. B. Pendry, *Phys. Rev. B* **69**, 241101 (2004).
16. J. Zhou *et al.*, *Phys. Rev. Lett.* **95**, 223902 (2005).
17. A. K. Popov, V. M. Shalaev, available at http://arxiv.org/abs/physics/0601055 (2006).
18. V. G. Veselago, *Sov. Phys. Usp.* **10**, 509 (1968).
19. M. Wegener, *Extreme Nonlinear Optics* (Springer, Berlin, 2004).
20. H. M. Barlow, *Nature* **173**, 41 (1954).
21. S.-Y. Chen, M. Maksimchuk, D. Umstadter, *Nature* **396**, 653 (1998).
22. Materials and Methods are available as supporting material on *Science* Online.
23. P. Guyot-Sionnest, W. Chen, Y. R. Shen, *Phys. Rev. B* **33**, 8254 (1986).
24. We thank the groups of S. W. Koch, J. V. Moloney, and C. M. Soukoulis for discussions. The research of M.W. is supported by the Leibniz award 2000 of the Deutsche Forschungsgemeinschaft (DFG), that of S.L. through a Helmholtz-Hochschul-Nachwuchsgruppe (VH-NG-232).

Supporting Online Material

www.sciencemag.org/cgi/content/full/313/5786/502/DC1
Materials and Methods
Figs. S1 and S2
References

26 April 2006; accepted 22 June 2006
10.1126/science.1129198

Reducing the Dimensionality of Data with Neural Networks

G. E. Hinton* and R. R. Salakhutdinov

High-dimensional data can be converted to low-dimensional codes by training a multilayer neural network with a small central layer to reconstruct high-dimensional input vectors. Gradient descent can be used for fine-tuning the weights in such "autoencoder" networks, but this works well only if the initial weights are close to a good solution. We describe an effective way of initializing the weights that allows deep autoencoder networks to learn low-dimensional codes that work much better than principal components analysis as a tool to reduce the dimensionality of data.

Dimensionality reduction facilitates the classification, visualization, communication, and storage of high-dimensional data. A simple and widely used method is principal components analysis (PCA), which finds the directions of greatest variance in the data set and represents each data point by its coordinates along each of these directions. We describe a nonlinear generalization of PCA that uses an adaptive, multilayer "encoder" network to transform the high-dimensional data into a low-dimensional code and a similar "decoder" network to recover the data from the code. Starting with random weights in the two networks, they can be trained together by minimizing the discrepancy between the original data and its reconstruction. The required gradients are easily obtained by using the chain rule to backpropagate error derivatives first through the decoder network and then through the encoder network (*1*). The whole system is called an "autoencoder" and is depicted in Fig. 1.

It is difficult to optimize the weights in nonlinear autoencoders that have multiple hidden layers (*2*–*4*). With large initial weights, autoencoders typically find poor local minima; with small initial weights, the gradients in the early layers are tiny, making it infeasible to train autoencoders with many hidden layers. If the initial weights are close to a good solution, gradient descent works well, but finding such initial weights requires a very different type of algorithm that learns one layer of features at a time. We introduce this "pretraining" procedure for binary data, generalize it to real-valued data, and show that it works well for a variety of data sets.

An ensemble of binary vectors (e.g., images) can be modeled using a two-layer network called a "restricted Boltzmann machine" (RBM) (*5*, *6*) in which stochastic, binary pixels are connected to stochastic, binary feature detectors using symmetrically weighted connections. The pixels correspond to "visible" units of the RBM because their states are observed; the feature detectors correspond to "hidden" units. A joint configuration (**v**, **h**) of the visible and hidden units has an energy (*7*) given by

$$E(\mathbf{v},\mathbf{h}) = -\sum_{i\in \text{pixels}} b_i v_i - \sum_{j\in \text{features}} b_j h_j - \sum_{i,j} v_i h_j w_{ij} \tag{1}$$

where v_i and h_j are the binary states of pixel *i* and feature *j*, b_i and b_j are their biases, and w_{ij} is the weight between them. The network assigns a probability to every possible image via this energy function, as explained in (*8*). The probability of a training image can be raised by

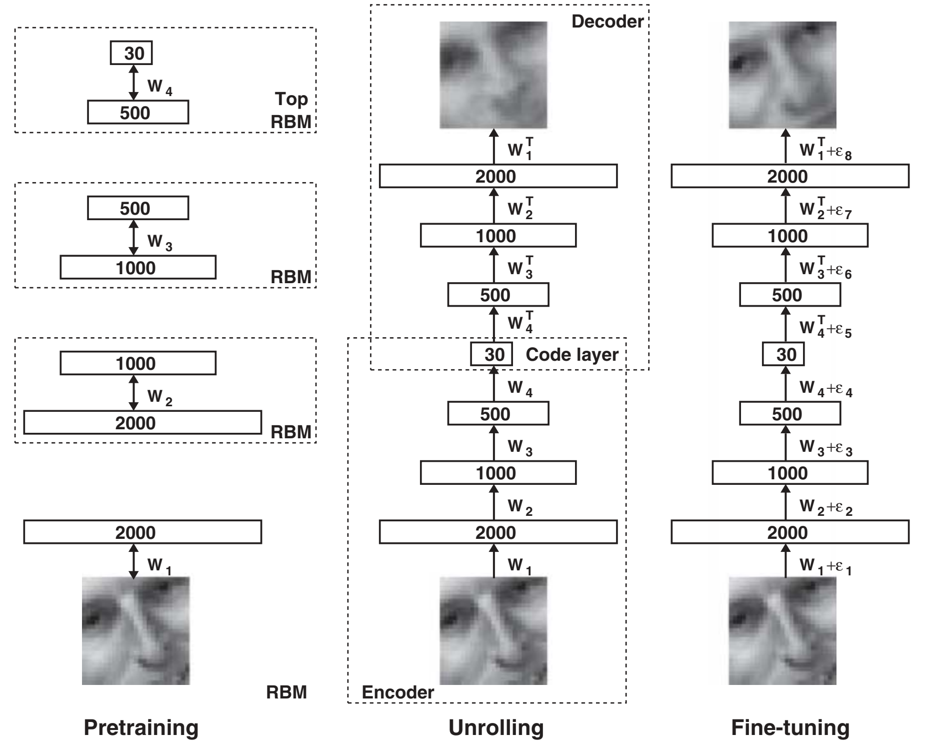


Fig. 1. Pretraining consists of learning a stack of restricted Boltzmann machines (RBMs), each having only one layer of feature detectors. The learned feature activations of one RBM are used as the "data" for training the next RBM in the stack. After the pretraining, the RBMs are "unrolled" to create a deep autoencoder, which is then fine-tuned using backpropagation of error derivatives.

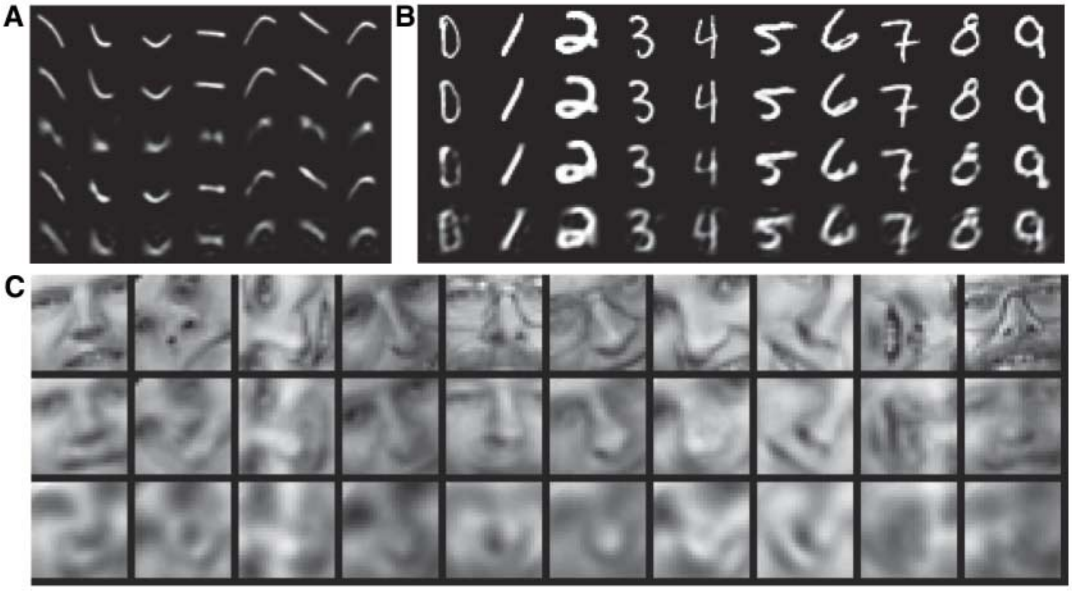


Fig. 2. (**A**) Top to bottom: Random samples of curves from the test data set; reconstructions produced by the six-dimensional deep autoencoder; reconstructions by "logistic PCA" (*8*) using six components; reconstructions by logistic PCA and standard PCA using 18 components. The average squared error per image for the last four rows is 1.44, 7.64, 2.45, 5.90. (**B**) Top to bottom: A random test image from each class; reconstructions by the 30-dimensional autoencoder; reconstructions by 30-dimensional logistic PCA and standard PCA. The average squared errors for the last three rows are 3.00, 8.01, and 13.87. (**C**) Top to bottom: Random samples from the test data set; reconstructions by the 30-dimensional autoencoder; reconstructions by 30-dimensional PCA. The average squared errors are 126 and 135.

Department of Computer Science, University of Toronto, 6 King's College Road, Toronto, Ontario M5S 3G4, Canada.

*To whom correspondence should be addressed; E-mail: hinton@cs.toronto.edu

adjusting the weights and biases to lower the energy of that image and to raise the energy of similar, “confabulated” images that the network would prefer to the real data. Given a training image, the binary state h_j of each feature detector j is set to 1 with probability $\sigma(b_j + \sum_i v_i w_{ij})$, where $\sigma(x)$ is the logistic function $1/[1 + \exp(-x)]$, b_j is the bias of j, v_i is the state of pixel i, and w_{ij} is the weight between i and j. Once binary states have been chosen for the hidden units, a “confabulation” is produced by setting each v_i to 1 with probability $\sigma(b_i + \sum_j h_j w_{ij})$, where b_i is the bias of i. The states of the hidden units are then updated once more so that they represent features of the confabulation. The change in a weight is given by

$$\Delta w_{ij} = \varepsilon\left(\langle v_i h_j\rangle_{\text{data}} - \langle v_i h_j\rangle_{\text{recon}}\right) \quad (2)$$

where ε is a learning rate, $\langle v_i h_j\rangle_{\text{data}}$ is the fraction of times that the pixel i and feature detector j are on together when the feature detectors are being driven by data, and $\langle v_i h_j\rangle_{\text{recon}}$ is the corresponding fraction for confabulations. A simplified version of the same learning rule is used for the biases. The learning works well even though it is not exactly following the gradient of the log probability of the training data (*6*).

A single layer of binary features is not the best way to model the structure in a set of images. After learning one layer of feature detectors, we can treat their activities—when they are being driven by the data—as data for learning a second layer of features. The first layer of feature detectors then become the visible units for learning the next RBM. This layer-by-layer learning can be repeated as many

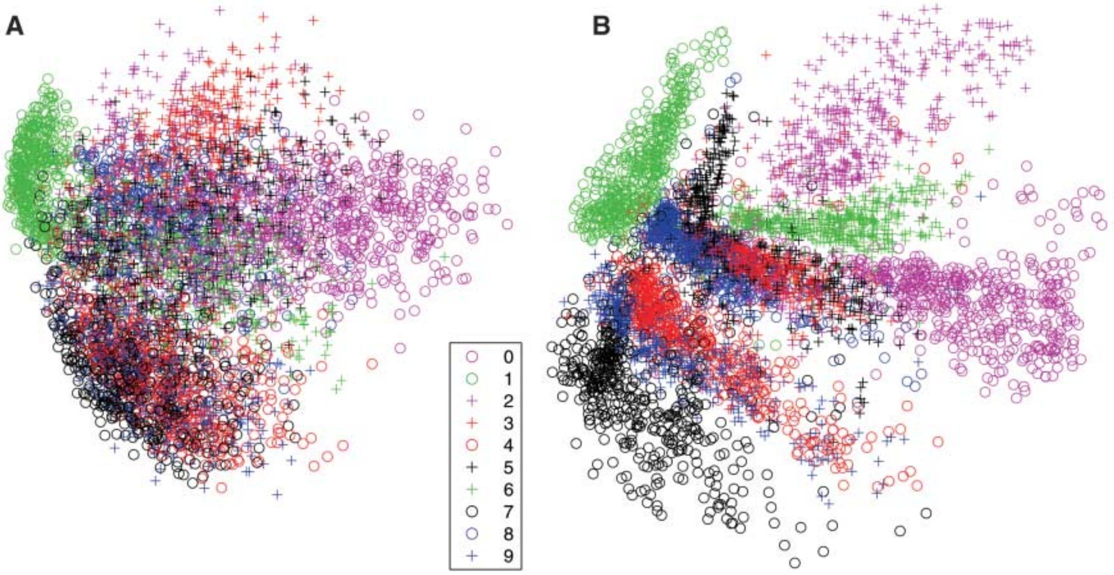


Fig. 3. (**A**) The two-dimensional codes for 500 digits of each class produced by taking the first two principal components of all 60,000 training images. (**B**) The two-dimensional codes found by a 784-1000-500-250-2 autoencoder. For an alternative visualization, see (*8*).

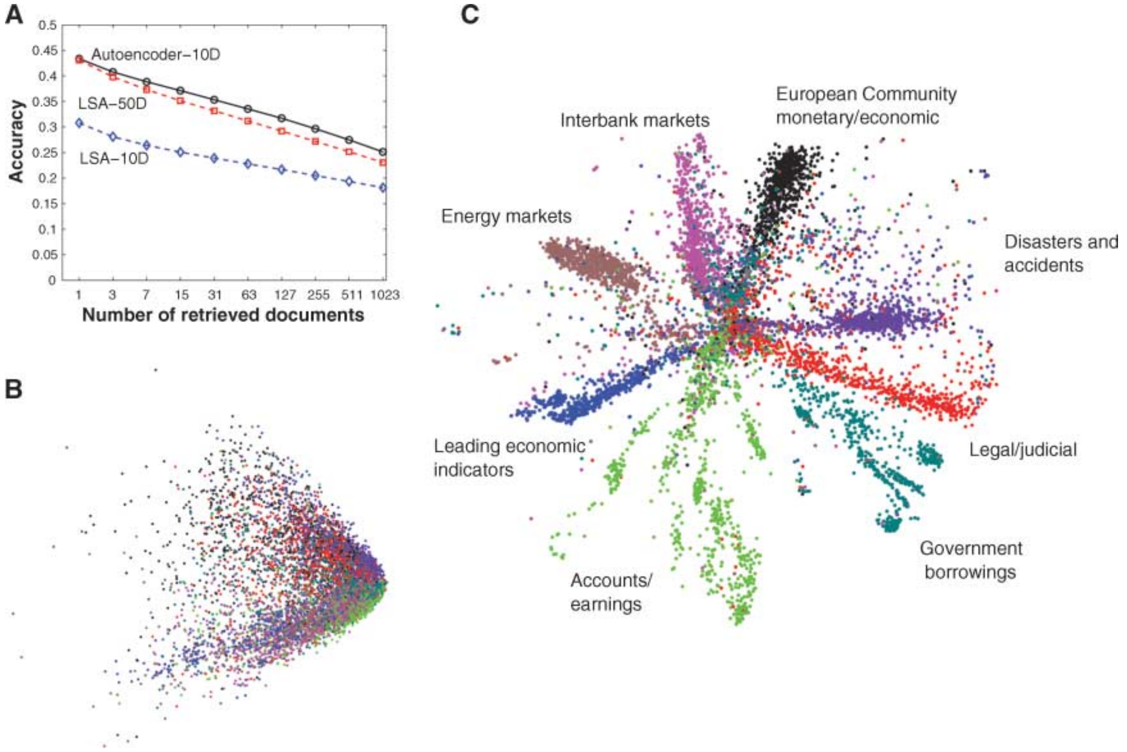


Fig. 4. (**A**) The fraction of retrieved documents in the same class as the query when a query document from the test set is used to retrieve other test set documents, averaged over all 402,207 possible queries. (**B**) The codes produced by two-dimensional LSA. (**C**) The codes produced by a 2000-500-250-125-2 autoencoder.

times as desired. It can be shown that adding an extra layer always improves a lower bound on the log probability that the model assigns to the training data, provided the number of feature detectors per layer does not decrease and their weights are initialized correctly (*9*). This bound does not apply when the higher layers have fewer feature detectors, but the layer-by-layer learning algorithm is nonetheless a very effective way to pretrain the weights of a deep autoencoder. Each layer of features captures strong, high-order correlations between the activities of units in the layer below. For a wide variety of data sets, this is an efficient way to progressively reveal low-dimensional, nonlinear structure.

After pretraining multiple layers of feature detectors, the model is "unfolded" (Fig. 1) to produce encoder and decoder networks that initially use the same weights. The global fine-tuning stage then replaces stochastic activities by deterministic, real-valued probabilities and uses backpropagation through the whole autoencoder to fine-tune the weights for optimal reconstruction.

For continuous data, the hidden units of the first-level RBM remain binary, but the visible units are replaced by linear units with Gaussian noise (*10*). If this noise has unit variance, the stochastic update rule for the hidden units remains the same and the update rule for visible unit i is to sample from a Gaussian with unit variance and mean $b_i + \sum_j h_j w_{ij}$.

In all our experiments, the visible units of every RBM had real-valued activities, which were in the range [0, 1] for logistic units. While training higher level RBMs, the visible units were set to the activation probabilities of the hidden units in the previous RBM, but the hidden units of every RBM except the top one had stochastic binary values. The hidden units of the top RBM had stochastic real-valued states drawn from a unit variance Gaussian whose mean was determined by the input from that RBM's logistic visible units. This allowed the low-dimensional codes to make good use of continuous variables and facilitated comparisons with PCA. Details of the pretraining and fine-tuning can be found in (*8*).

To demonstrate that our pretraining algorithm allows us to fine-tune deep networks efficiently, we trained a very deep autoencoder on a synthetic data set containing images of "curves" that were generated from three randomly chosen points in two dimensions (*8*). For this data set, the true intrinsic dimensionality is known, and the relationship between the pixel intensities and the six numbers used to generate them is highly nonlinear. The pixel intensities lie between 0 and 1 and are very non-Gaussian, so we used logistic output units in the autoencoder, and the fine-tuning stage of the learning minimized the cross-entropy error $[-\sum_i p_i \log \hat{p}_i - \sum_i (1-p_i) \log(1-\hat{p}_i)]$, where p_i is the intensity of pixel i and $\hat{p}_i$ is the intensity of its reconstruction.

The autoencoder consisted of an encoder with layers of size (28 × 28)-400-200-100-50-25-6 and a symmetric decoder. The six units in the code layer were linear and all the other units were logistic. The network was trained on 20,000 images and tested on 10,000 new images. The autoencoder discovered how to convert each 784-pixel image into six real numbers that allow almost perfect reconstruction (Fig. 2A). PCA gave much worse reconstructions. Without pretraining, the very deep autoencoder always reconstructs the average of the training data, even after prolonged fine-tuning (*8*). Shallower autoencoders with a single hidden layer between the data and the code can learn without pretraining, but pretraining greatly reduces their total training time (*8*). When the number of parameters is the same, deep autoencoders can produce lower reconstruction errors on test data than shallow ones, but this advantage disappears as the number of parameters increases (*8*).

Next, we used a 784-1000-500-250-30 autoencoder to extract codes for all the handwritten digits in the MNIST training set (*11*). The Matlab code that we used for the pretraining and fine-tuning is available in (*8*). Again, all units were logistic except for the 30 linear units in the code layer. After fine-tuning on all 60,000 training images, the autoencoder was tested on 10,000 new images and produced much better reconstructions than did PCA (Fig. 2B). A two-dimensional autoencoder produced a better visualization of the data than did the first two principal components (Fig. 3).

We also used a 625-2000-1000-500-30 autoencoder with linear input units to discover 30-dimensional codes for grayscale image patches that were derived from the Olivetti face data set (*12*). The autoencoder clearly outperformed PCA (Fig. 2C).

When trained on documents, autoencoders produce codes that allow fast retrieval. We represented each of 804,414 newswire stories (*13*) as a vector of document-specific probabilities of the 2000 commonest word stems, and we trained a 2000-500-250-125-10 autoencoder on half of the stories with the use of the multiclass cross-entropy error function $[-\sum_i p_i \log \hat{p}_i]$ for the fine-tuning. The 10 code units were linear and the remaining hidden units were logistic. When the cosine of the angle between two codes was used to measure similarity, the autoencoder clearly outperformed latent semantic analysis (LSA) (*14*), a well-known document retrieval method based on PCA (Fig. 4). Autoencoders (*8*) also outperform local linear embedding, a recent nonlinear dimensionality reduction algorithm (*15*).

Layer-by-layer pretraining can also be used for classification and regression. On a widely used version of the MNIST handwritten digit recognition task, the best reported error rates are 1.6% for randomly initialized backpropagation and 1.4% for support vector machines. After layer-by-layer pretraining in a 784-500-500-2000-10 network, backpropagation using steepest descent and a small learning rate achieves 1.2% (*8*). Pretraining helps generalization because it ensures that most of the information in the weights comes from modeling the images. The very limited information in the labels is used only to slightly adjust the weights found by pretraining.

It has been obvious since the 1980s that backpropagation through deep autoencoders would be very effective for nonlinear dimensionality reduction, provided that computers were fast enough, data sets were big enough, and the initial weights were close enough to a good solution. All three conditions are now satisfied. Unlike nonparametric methods (*15*, *16*), autoencoders give mappings in both directions between the data and code spaces, and they can be applied to very large data sets because both the pretraining and the fine-tuning scale linearly in time and space with the number of training cases.

References and Notes

1. D. C. Plaut, G. E. Hinton, *Comput. Speech Lang.* **2**, 35 (1987).
2. D. DeMers, G. Cottrell, *Advances in Neural Information Processing Systems 5* (Morgan Kaufmann, San Mateo, CA, 1993), pp. 580–587.
3. R. Hecht-Nielsen, *Science* **269**, 1860 (1995).
4. N. Kambhatla, T. Leen, *Neural Comput.* **9**, 1493 (1997).
5. P. Smolensky, *Parallel Distributed Processing: Volume 1: Foundations*, D. E. Rumelhart, J. L. McClelland, Eds. (MIT Press, Cambridge, 1986), pp. 194–281.
6. G. E. Hinton, *Neural Comput.* **14**, 1711 (2002).
7. J. J. Hopfield, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2554 (1982).
8. See supporting material on *Science* Online.
9. G. E. Hinton, S. Osindero, Y. W. Teh, *Neural Comput.* **18**, 1527 (2006).
10. M. Welling, M. Rosen-Zvi, G. Hinton, *Advances in Neural Information Processing Systems 17* (MIT Press, Cambridge, MA, 2005), pp. 1481–1488.
11. The MNIST data set is available at http://yann.lecun.com/exdb/mnist/index.html.
12. The Olivetti face data set is available at www.cs.toronto.edu/ roweis/data.html.
13. The Reuter Corpus Volume 2 is available at http://trec.nist.gov/data/reuters/reuters.html.
14. S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, R. A. Harshman, *J. Am. Soc. Inf. Sci.* **41**, 391 (1990).
15. S. T. Roweis, L. K. Saul, *Science* **290**, 2323 (2000).
16. J. A. Tenenbaum, V. J. de Silva, J. C. Langford, *Science* **290**, 2319 (2000).

17. We thank D. Rumelhart, M. Welling, S. Osindero, and S. Roweis for helpful discussions, and the Natural Sciences and Engineering Research Council of Canada for funding. G.E.H. is a fellow of the Canadian Institute for Advanced Research.

Supporting Online Material

www.sciencemag.org/cgi/content/full/313/5786/504/DC1
Materials and Methods
Figs. S1 to S5
Matlab Code

20 March 2006; accepted 1 June 2006
10.1126/science.1127647

Early Pleistocene Glacial Cycles and the Integrated Summer Insolation Forcing

Peter Huybers

Long-term variations in Northern Hemisphere summer insolation are generally thought to control glaciation. But the intensity of summer insolation is primarily controlled by 20,000-year cycles in the precession of the equinoxes, whereas early Pleistocene glacial cycles occur at 40,000-year intervals, matching the period of changes in Earth's obliquity. The resolution of this 40,000-year problem is that glaciers are sensitive to insolation integrated over the duration of the summer. The integrated summer insolation is primarily controlled by obliquity and not precession because, by Kepler's second law, the duration of the summer is inversely proportional to Earth's distance from the Sun.

A link between changes in glacial extent and Earth's orbital configuration was apparently first proposed by Adhémar (*1*, *2*), who postulated that the Antarctic ice sheet exists because the Southern Hemisphere winter is 8 days longer than the Northern one. In this case, winter is the period between the equinoxes. This difference in duration follows from Kepler's second law and from the fact that Earth's closest approach to the Sun, that is, perihelion, currently occurs during Northern Hemisphere winter. Croll modified this hypothesis, arguing that the decrease in insolation associated with being further from the Sun leads to glaciation (*2*, *3*). Milankovitch, in turn, argued that summer insolation determines glaciation (*4*). More recently, once proxies of past glaciation showed that late Pleistocene glacial cycles occurred at ~100,000-year (100-ky) intervals (*5*), the amplitude envelope of the precession (i.e., the eccentricity) was identified as accounting for the 100-ky glacial cycles (*5*–*7*).

This thread of glacial hypotheses thus implies that precession of the equinoxes controls the occurrence of glacial cycles. Indeed, variations in the intensity of summer insolation are primarily controlled by precession. For example, average insolation on the 21st day of June at 65°N has 80% of its variance at the precession periods (1/21 ky ± 1/100 ky). The caloric summer half-year at 65°N, defined as the energy received during the half of the year with the greatest insolation intensity (*4*), also has more than half its variance in the precession bands. But a major problem exists for the standard orbital hypothesis of glaciation: Late Pliocene and early Pleistocene glacial cycles occur at intervals of 40 ky (*8*–*11*), matching the obliquity period, but have negligible 20-ky variability.

One possibility is that the latitudinal gradient in insolation, which enhances obliquity over precession, is more important than local insolation (*11*). However, models used to explore the effects of changes in the insolation gradient have found that local insolation is the more important control on glacial mass balance (*12*). Simple models that used summer insolation as the forcing (*13*–*15*) exhibited more precession-period variability than is observed in the early Pleistocene climate record. Another possibility is that glaciation is controlled by the annual average insolation, which is independent of precession, but this hypothesis requires glacial mass balance to be equally sensitive to winter and summer insolation (*16*). One climate model (*17*) that is forced by the complete seasonal cycle showed predominantly precession-period glacial variability during the early Pleistocene, whereas another, more sophisticated, coupled climate–ice sheet model (*18*) showed primarily obliquity period variability (although the latter model is for Antarctica near ~34 My ago), and neither study identified mechanisms for the differing sensitivities to orbital variations. The origins of strong obliquity over precession-period glacial variability during the early Pleistocene remain unresolved.

Tying insolation at the top of the atmosphere to climate on the ground poses a serious challenge. It is useful to consider empirical relationships between insolation (*19*) and modern temperature (*20*). Insolation lagged by 30 days shows an excellent correlation with zonally and diurnally averaged land temperature, $\overline{T}$, for latitude bands north of 30°N ($r^2 > 0.99$) (Fig. 1C). Insolation is apparently a good predictor of $\overline{T}$.

A more complicated relationship might have been expected between insolation and $\overline{T}$ when one considers processes such as reflection of radiation by snow, ice, and clouds; changes in heat storage; and the redistribution of heat by the ocean and the atmosphere. The linear relationship between insolation and average temperature does not exclude the importance of these processes but does suggest that their aggregate influence is also correlated with the insolation. Furthermore, the combined heat

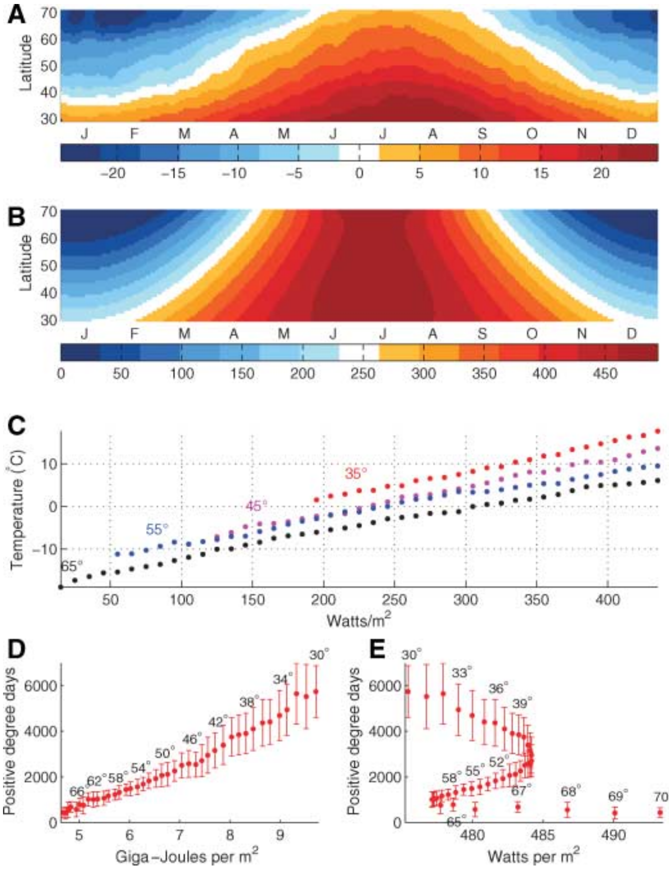


Fig. 1. Relationships between insolation and temperatures. (**A**) Temperature in °C contoured as a function of latitude and month. Temperatures, $\overline{T}$, are diurnal averages from WMO stations and are averaged according to latitude after adjusting for elevation using a lapse rate of 6.5°C/km. (**B**) Insolation at the top of the atmosphere. (**C**) $\overline{T}$ plotted against insolation for different latitudes ($r^2 > 0.99$). Latitude bins are 10°, and insolation bins are 10 W/m^2 where insolation has been lagged by 1 month. (**D**) Positive degree days plotted against summer energy ($r^2 = 0.98$). (**E**) Positive degree days plotted against the intensity of diurnally averaged insolation on June 21st ($r^2 = 0.04$).

Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138, USA. E-mail: phuybers@fas.harvard.edu

transport of the ocean and the atmosphere to latitudes above 30°N amounts to 5 PW (petaWatts) (*21*) and, when spatially averaged, corresponds to 40 W/m^2, or less than 10% of the summer insolation at the top of the atmosphere at any latitude. In this light, it is reasonable for insolation to primarily control local temperature, particularly during the summer months.

If one accepts the empirical relationship between insolation and temperature, then what is the best measure of insolation's influence on ablation? It is not mean annual insolation: The ablation season is not more than 6 months in duration, and the temperature during the rest of the year seems largely irrelevant (*22*). Mean summer insolation is a more likely candidate. However, defining summer is difficult because the length of the ablation season should depend on the insolation cycle itself as well as other environmental factors.

A good measure of air temperature's influence on annual ablation is the sum of positive degree days (*22*, *23*), defined as $S = \sum_i \alpha_i T_i$, where T_i is mean daily temperature on day i and α is one when $T_i \geq 0°C$ and zero otherwise. A quantity analogous to S can be defined for insolation. For latitudes between 40° to 70°N, the temperature is near 0°C when insolation intensity is between 250 and 300 W/m^2 (Fig. 1C), and τ = 275 W/m^2 is taken as a threshold (*24*). The number of degree days is postulated to follow the sum of the insolation on days exceeding this threshold, $J = \sum_i \beta_i (W_i \times 86{,}400)$, where J is termed the summer energy and is measured in joules. W_i is mean insolation in W/m^2 on day i, and β equals one when $W_i \geq \tau$ and zero otherwise. Note that ablation responds to both radiative transfer and heat flux from the atmosphere into the ice, but this distinction is not made because insolation and temperature are strongly correlated.

S, computed by using $\overline{T}$, monotonically decreases from 6000 at 30°N to 400 at 70°N. The summer energy also steadily decreases toward high latitudes and is highly correlated with the positive degree days (r^2 = 0.98) (Fig. 1D). In contrast, the average insolation intensity on June 21st has a more complicated dependence on latitude (owing to the tradeoff between zenith angle and hours of daylight) and has a low correlation with the positive degree days (r^2 = 0.04) (Fig. 1E). It is perhaps unsurprising that insolation on June 21st fails to correspond to positive degree days. For similar reasons, one would not expect temperature on a single day of the year to adequately predict annual ablation.

Long-term variations in the duration of the summertime and intensity of summer insolation are primarily controlled by the precession of the equinoxes, with more than 80% of their respective variances within 1/21 ky ± 1/100 ky (Fig. 2, A and B). Duration and intensity are, however, anticorrelated. This is the Achilles' heel of precession control of glaciation: just when Earth is closest to the sun during summer, summertime is shortest. When the intensity is integrated over the summertime, precession-related changes in duration and intensity nearly balance one another (*25*), and the obliquity component is dominant. When τ = 275 W/m^2, 80% of the summer energy variance is in the obliquity band (1/41 ky ± 1/100 ky) (Fig. 2, C and D).

As an example, Earth's orbital configuration when perihelion occurs variously at the equinoxes and at the solstices is shown (Fig. 3) for the interval between 220 and 200 ky ago. When perihelion occurs at summer solstice rather than winter solstice, mean summer insolation at 65°N is 54 W/m^2 greater (assuming a fixed obliquity of 23.3°), but summer is also 13 days shorter. Changes in the orientation of perihelion with respect to the seasons cause deviations of no more than ±0.1 GJ (giga-Joules) from a mean summer energy of 5.0 GJ. In contrast, if perihelion is fixed at summer solstice, an increase in obliquity from 22.1° to 24.5° results in an average increase in summer intensity of 24 W/m^2 (Fig. 3C), an increase in summer duration from 133 to 137 days, and an increase in summer energy from 4.9 to 5.3 GJ/m^2 (*26*).

Changes in accumulation, although more difficult to infer from insolation, may also contribute to changes in the glacial mass

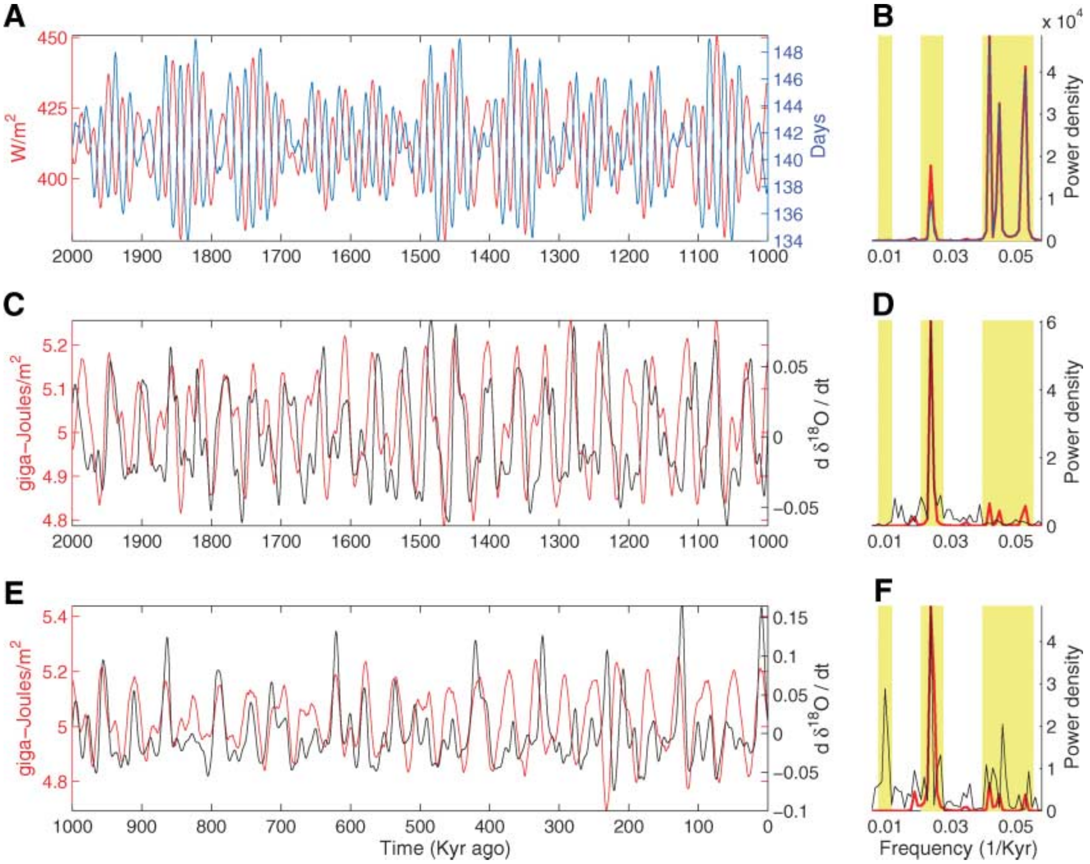


Fig. 2. Insolation forcing and Pleistocene glacial variability. (**A**) Number of days that insolation is above 275 W/m^2 (blue) and the average insolation intensity during this interval (red). Intensity and duration are anticorrelated. (**B**) Spectral estimate of the duration (blue) and intensity (red), showing that the majority of the variability is at the precession periods. Shaded bands from left to right indicate the 100-ky, 41-ky (obliquity), and 21-ky (precession) bands. (**C**) Summer energy (red) and the time rate of change of $\delta^{18}O$ (black) for the early Pleistocene and (**D**) the corresponding spectral estimates. Positive rates of change indicate decreasing ice volume. Variability in both records is predominantly at the 41-ky obliquity period. (**E** and **F**) Same as (C) and (D) but for the late Pleistocene. The time rate of change of $\delta^{18}O$ has variability at the 100-ky period not present in the forcing.

balance. In addition to high-latitude summer energy increasing when obliquity is large, winter energy decreases, possibly decreasing winter temperature and causing a decrease in atmospheric moisture and glacial accumulation.

So far, only modern observations have been used to argue that summer energy is a better indicator of glacial variability than insolation intensity. It remains to test this result against past glacial variations. Changes in summer energy are expected to correspond to rates of ablation and thus are most directly compared against rates of ice volume change (*27*). After smoothing using an 11-ky tapered window, the time derivative of a composite δ^{18}O record is used as a proxy for ice volume change (*28*). Importantly, the age model for the proxy record does not rely upon orbital assumptions.

There is an excellent correspondence between summer energy at 65°N and the rate of ice volume change. For the early Pleistocene, 70% of the variance in the rate of ice volume change is concentrated at the obliquity band (1/41 ky ± 1/100 ky, $P = 0.01$) (Fig. 2, C and D), and the obliquity band is in phase and 80% coherent with the summer energy ($P = 0.01$). There is also a significant correlation between the amplitude of the summer energy forcing and the amplitude of ablation ($r^2 = 0.5$, $P = 0.01$), whereas June 21st insolation shows negligible correlation ($r^2 = 0.1$) (*29*). Parametrization of the insolation forcing using summer energy seems to resolve the question of why early Pleistocene glacial cycles occur primarily at 40-ky intervals. More generally, summer energy may explain why obliquity appears to be the primary period of glacial variability throughout the glaciated portions of the Cenozoic (*30*).

The concept of summer energy also has implications for the ~100-ky glacial variability during the late Pleistocene (*31*). Obliquity period variability remains the strongest component of ice volume change during the late Pleistocene, having nearly the same magnitude as during the early Pleistocene but accounting for a smaller fraction of variance (40%) because of enhanced precession (26% at 1/21 ky ± 1/100 ky) and 100-ky period variance (22% at 1/100 ky ± 1/300 ky) (Fig. 2, E and F). Note that the rate of change used here, rather than magnitude of ice volume, has relatively more variance at high frequencies.

The amplitude of the summer energy and rates of ablation show less agreement during the late Pleistocene ($r^2 = 0.4$) than during the early Pleistocene. The most rapid ablation events, known as terminations, follow periods of greatest ice volume (*32*), suggesting that the sensitivity to summer energy depends on the amount of ice volume. To quantify this effect, the amount of ice volume is estimated with the use of δ^{18}O 10 ky before peak ablation, and sensitivity is defined as the ratio between the amplitude of ablation and the amplitude of the local maximum in summer energy nearest in time. A significant correlation is observed between ice volume and sensitivity ($r^2 = 0.6$). Perhaps large ice sheets are inherently more unstable (*13*), or perhaps they are more strongly forced by local insolation because they extend to lower latitudes.

A cooling climate during the Pleistocene (*30*, *33*) may have permitted ice volume to build up over multiple forcing cycles, allowing sensitivity to increase until an increase in summer energy triggers a glacial termination. In agreement with earlier results (*16*), terminations occur at intervals of about two (80-ky) or three (120-ky) obliquity cycles, on average giving the ~100-ky variability. A cooling Pleistocene climate may also be expected to increase the threshold τ, at which melting occurs. A higher τ makes summer energy more variable and more sensitive to precession variations (fig. S1). For example, raising τ from 275 to 340 W/m^2 more than doubles the summer energy variance and gives equal precession and obliquity period variability. Thus, a cooling climate and increased τ may help explain both the larger glacial variations and the appearance of precession period variability during the late Pleistocene.

The hypothesis presented here follows from both Adhémar's argument regarding seasonal duration and Croll and Milankovitch's argument regarding insolation intensity. Taking duration and intensity together, it now appears that summer energy controls early Pleistocene glacial variability. However, the 100-ky glacial cycles of the late Pleistocene have a more complicated relationship with the forcing, and their explanation will require a better understanding of ice sheet–climate interactions.

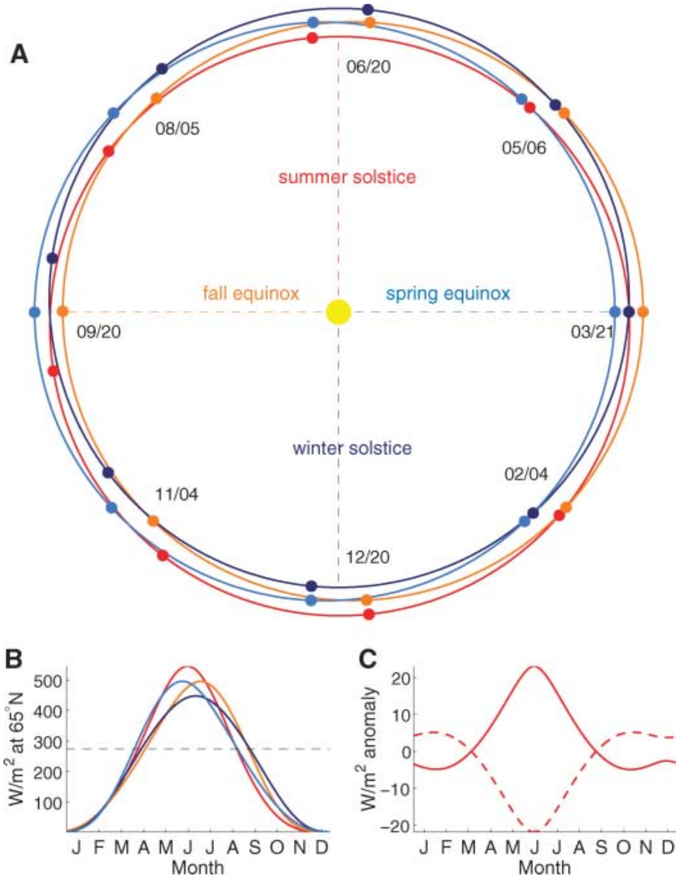


Fig. 3. The Earth's variable orbit around the Sun. (**A**) Earth's orbit when perihelion occurs at Northern Hemisphere summer solstice (red), fall equinox (orange), winter solstice (blue), and spring equinox (light blue), corresponding to the orbital configurations near 220.2, 214.6, 209.2, and 203.5 ky ago, respectively. The eccentricity of Earth's orbit averages 0.05 during this interval. The orbit is to scale and oriented so that spring equinox always occurs at the three o'clock position. March 21st is referenced to the spring equinox, and the location of the Earth is shown every 45.7 days (colored dots with dates given as month/day). Earth moves counterclockwise. The orbit having perihelion during Northern Hemisphere summer (red) reaches fall equinox the soonest. (**B**) Seasonal variations in insolation at 65°N. The *x* axis is labeled with the midpoint of each month. The orbit with perihelion at summer solstice (red) achieves the greatest insolation intensity but also has the shortest duration above a 275 W/m^2 threshold (indicated by the horizontal dashed line). (**C**) Anomalies in insolation for obliquity values of 22.1° (dashed) and 24.5° (solid) relative to a mean obliquity of 23.3° for the orbit with perihelion at summer solstice.

References and Notes

1. J. Adhémar, *Révolutions de la Mer* (Carilian-Goeury et V. Dalmont, Paris, 1842).
2. E. Bard, *C. R. Geosci.* **336**, 603 (2004).
3. J. Croll, *Philos. Mag.* **28**, 121 (1864).
4. M. Milankovitch, *Kanon der Erdbestrahlung und seine Andwendung auf das Eiszeitenproblem* (Royal Serbian Academy, Belgrade, 1941).
5. J. D. Hays, J. Imbrie, N. J. Shackleton, *Science* **194**, 1121 (1976).
6. J. Imbrie, J. Z. Imbrie, *Science* **207**, 943 (1980).
7. J. Imbrie *et al.*, *Paleoceanography* **8**, 699 (1993).
8. J. Imbrie *et al.*, *Paleoceanography* **7**, 701 (1992).
9. R. Tiedemann, M. Sarnthein, N. J. Shackleton, *Paleoceanography* **9**, 619 (1994).

10. L. Lisiecki, M. Raymo, *Paleoceanography* **20**, 10.1029/2004PA001071 (2005).
11. M. Raymo, K. Nisancioglu, *Paleoceanography* **18**, 10.1029/2002PA000791 (2003).
12. K. Nisancioglu, thesis, Massachusetts Institute of Technology (2004).
13. P. Clark, R. Alley, D. Pollard, *Science* **286**, 1104 (1999).
14. E. Tziperman, H. Gildor, *Paleoceanography* **18**, 10.1029/2001PA000627 (2003).
15. D. Paillard, *Nature* **391**, 378 (1998).
16. P. Huybers, C. Wunsch, *Nature* **434**, 491 (2005).
17. A. Berger, X. Li, M. Loutre, *Quat. Sci. Rev.* **18**, 1 (1999).
18. R. DeConto, D. Pollard, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **198**, 39 (2003).
19. A. Berger, M. F. Loutre, *Earth Planet. Sci. Lett.* **111**, 369 (1992).
20. Daily average surface temperatures are estimated by using the network of 8892 World Meteorological Organization (WMO) stations above 30°N for the years 1994 to 1999. All stations that have greater than 80% data coverage are used. Data gaps are filled by linear interpolation. Stations are standardized to 1 km of elevation assuming a lapse rate of 6.5°C/km, were binned according to 1° or 10° latitude bands (as indicated in the text), and are then averaged together. Lastly, each of the six consecutive seasonal cycles are averaged together, yielding average annual cycles as a function of latitude.
21. C. Wunsch, *J. Clim.* **18**, 4374 (2005).
22. W. Paterson, *Physics of Glaciers* (Pergamon Press, Oxford, ed. 3, 1994).
23. R. Braithwaite, Y. Zhang, *J. Glaciol.* **152**, 7 (2000).
24. The use of a constant value for τ illustrates the concept of summer energy. A more detailed description would take into account that τ is expected to be spatially and temporally variable, depending on factors such as elevation, albedo, clouds, heat transport, and greenhouse gas concentrations. Note, however, that results are not sensitive to plausible choices of τ and that values less than 325 W/m^2 yield broadly consistent summer energies (fig. S1). Summer energy values at 65°N are given in table S1.
25. The relationship between insolation intensity and insolation energy is more precisely illustrated by noting that $I \propto 1/r^2$, where I is insolation intensity and r is the distance from the Earth to the Sun. Conservation of angular momentum (or, equivalently, Kepler's second law) dictates that $dt \propto r^2 d\lambda$, where dt is an infinitesimal change in time and $d\lambda$ the corresponding change in solar longitude. The energy received by the Earth is then $J = Idt \propto d\lambda$. In contrast with I, the J between any two solar longitudes is independent of r and, thus, independent of the precession of the equinoxes.
26. Are past changes in summer energy sufficient to cause the waxing and waning of ice sheets? Although a full answer requires a realistic model of Pleistocene climate, some indication is provided by modern glacial changes: A 2.4 W/m^2 global average increase in radiative forcing caused by greenhouse gases (*34*) has apparently led to a general decrease in glacial mass (*35*), suggesting that glaciers are sensitive to relatively small changes in the radiation budget.
27. C. Wunsch, *Clim. Dyn.* **20**, 353 (2003).
28. Materials and Methods are available as supporting material on *Science* Online.
29. Amplitude cross correlation was computed by pairing local maxima in insolation with the nearest (in time) maximum in the rate of change of ice volume. Before identifying maxima, both the δ^{18}O record and the summer energy were smoothed by using an 11-ky tapered window. There are 34 local maxima in summer energy at 65°N between 2 and 1 My ago and another 34 between 1 My ago and the present. Squared cross correlations of 0.4 and higher have P values of less than 0.01. Spectral and coherence analysis is performed by using Thomson's multitaper method (*36*).
30. J. C. Zachos, N. J. Shackleton, J. S. Revenaugh, H. Pälike, B. P. Flower, *Science* **292**, 274 (2001).
31. Similar with the early Pleistocene, late Pleistocene ice volume change has significant variability concentrated at the obliquity band, which is in phase and coherent with summer energy ($P = 0.01$). That the obliquity component of summer energy varies symmetrically between the hemispheres helps explain the symmetry of glacial variations between the hemispheres. Also, the increase in summer energy near 420 ky ago, absent in measures of summer insolation forcing relying on intensity, helps explain the corresponding glacial termination.
32. M. E. Raymo, *Paleoceanography* **12**, 577 (1997).
33. M. Raymo, *Annu. Rev. Earth Planet. Sci.* **22**, 353 (1994).
34. J. Houghton *et al.*, Eds., *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, New York, 2001).
35. J. Oerlemans, *Science* **308**, 675 (2005); published online 3 March 2005 (10.1126/science.1107046).
36. D. Percival, A. Walden, *Spectral Analysis for Physical Applications* (Cambridge Univ. Press, Cambridge, 1993).

37. This paper benefited from discussion with E. Boyle, B. Curry, M. Raymo, P. Stone, E. Tziperman, and C. Wunsch. J. Levine provided valuable assistance in calculating the insolation. The NSF paleoclimate program supported this work under grant no. ATM-0455470.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1125249/DC1
SOM Text
Figs. S1 to S3
References

23 January 2006; accepted 9 June 2006
Published online 22 June 2006;
10.1126/science.1125249
Include this information when citing this paper.

A Thick Cloud of Neptune Trojans and Their Colors

Scott S. Sheppard[1]* and Chadwick A. Trujillo[2]

The dynamical and physical properties of asteroids offer one of the few constraints on the formation, evolution, and migration of the giant planets. Trojan asteroids share a planet's semimajor axis but lead or follow it by about 60° near the two triangular Lagrangian points of gravitational equilibrium. Here we report the discovery of a high-inclination Neptune Trojan, 2005 TN$_{53}$. This discovery demonstrates that the Neptune Trojan population occupies a thick disk, which is indicative of "freeze-in" capture instead of in situ or collisional formation. The Neptune Trojans appear to have a population that is several times larger than the Jupiter Trojans. Our color measurements show that Neptune Trojans have statistically indistinguishable slightly red colors, which suggests that they had a common formation and evolutionary history and are distinct from the classical Kuiper Belt objects.

The Neptune Trojans are only the fourth observed stable reservoir of small bodies in our solar system; the others are the Kuiper Belt, main asteroid belt, and jovian Trojans. The Trojan reservoirs of the giant planets lie between the rocky main belt asteroids and the volatile-rich Kuiper Belt. The effects of nebular gas drag (*1*), collisions (*2*), planetary migration (*3*, *4*), overlapping resonances (*5*, *6*), and the mass growth of the planets (*7*, *8*) all potentially influence the formation and evolution of the Neptune Trojans. The number of Jupiter Trojans is comparable to the main asteroid belt (*9*). One Neptune Trojan was discovered serendipitously in 2001 (*10*). Our ongoing dedicated Trojan survey has found three additional Neptune Trojans (Table 1).

Stable minor planets in the triangular Lagrangian Trojan regions, called the leading L4 and trailing L5 points, are said to be in a 1:1 resonance with the planet because each completes one orbit about the Sun with the period of the parent planet. The Neptune Trojans are distinctly different from other known Neptune resonance populations found in the Kuiper Belt. Kuiper Belt resonances such as the 3:2 (which Pluto occupies) and 2:1 may owe their existence to sweeping resonance capture of the migrating planets (*11*). The Neptune Trojans, however, would be lost because of migration and are not captured during this process (*3*, *4*, *10*).

Numerical dynamical stability simulations have shown that Neptune may retain up to 50% of its original Trojan population over the age of the solar system after any marked planetary migration (*4*, *12*). These simulations also demonstrate that Saturn and Uranus are not expected to have any substantial primordial Trojan populations. Recent numerical simulations of small bodies temporarily passing through the giant planet region, such as Centaurs, have shown that Neptune cannot currently efficiently capture Trojans even for short periods of time (*4*, *13*). Thus, capture or formation of the Trojans at the Lagrangian regions likely occurred during or just after the planet formation epoch, when conditions in the solar system were vastly different from those now. We numerically integrated (*14*) several orbits similar to each of the known Neptune Trojans and found that the majority of test particles near each known Neptune Trojan were stable over the age of the solar system.

Various mechanisms have been proposed that dissipated asteroid orbital energy to perma-

[1] Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road NW, Washington, DC 20015, USA. [2] Gemini Observatory, 670 North A'ohoku Place, Hilo, HI 96720, USA.

*To whom correspondence should be addressed. E-mail: sheppard@dtm.ciw.edu

nently capture bodies in the Lagrangian regions of the planets. Neptune's formation was probably quite different from Jupiter's (*15*), and thus gas drag (*1*) or rapid mass growth of the planet (*2*, *7*), as suggested for Jupiter Trojan capture, was probably not effective near Neptune. This suggests that the formation and possible capture of Neptune's Trojans was likely independent of the planet formation process. One possible mechanism is some type of "freeze-in" capture (*6*). This may occur if the orbits of the giant planets become excited and perturb many of the small bodies throughout the solar system. Once the orbits of the planets stabilize, any chance objects in the Lagrangian Trojan regions become stable and thus are trapped. A second mechanism proposed is collisional interactions within the Lagrangian region (*2*, *16*), and a third possible mechanism is in situ accretion of the Neptune Trojans from a subdisk of debris formed from postmigration collisions (*2*). The two collisional theories above predict low inclination Trojans, whereas freeze-in allows for high-inclination bodies.

Through our survey using the Magellan-Baade 6.5-m telescope, we have discovered one high-inclination (2005 TN_{53}; inclination $i \sim 25$ degrees) and two low-inclination (2004 UP_{10} and 2005 TO_{74}; $i < 5°$) Trojans (Table 1). Because we have surveyed only in the low-latitude Neptune L4 region (within 1.5° of Neptune's orbit and the ecliptic), we have been heavily biased toward finding low-inclination Trojans. We can correct for this bias by determining the probability of detecting high-inclination Trojans in our low-inclination fields. A high-inclination Trojan on a nearly circular orbit like 2005 TN_{53} would only spend about 2% of its orbit within the 1.5° of ecliptic latitude that our survey covered. Thus, for every high-inclination Trojan discovered by our survey, we should expect tens of high-inclination Trojans (i.e., 50^{+75}_{-35}) outside this latitude range, assuming Poisson statistics. The low-inclination Trojans 2004 UP_{10} ($i \sim 1.4°$) and 2005 TO_{74} ($i \sim 5°$), respectively, spend about 50% and 10% of their orbits within our survey latitudes. Thus, the ratio of high- to low-inclination L4 Trojans is 50^{+75}_{-35}:12^{+10}_{-7}, or about 4:1. Statistically, there may be at least as many, although most likely more, high-inclination Trojans as there are low-inclination Trojans. Our discovery of 2005 TN_{53} indicates that the Neptune Trojan population occupies a thick disk, like the Jupiter Trojans. The collisional and in situ accretion Neptune Trojan formation models predict few, if any, high-inclination Trojans. Unless there has been recent excitation in the Neptune L4 region, freeze-in capture (*6*) or some variation on it (*17*) appears to be the most likely capture process.

We further explored the physical properties of the Neptune Trojans by obtaining optical colors of each Trojan (Table 2). We found that the four known Neptune Trojans have indistinguishable optical colors, which are consistent with a common formation and evolution history for all. This is unlike the Kuiper Belt objects (KBOs), which show a large color diversity from gray to some of the reddest objects in the solar system (*18*). We found that the Neptune Trojans have optical colors a little redder than the gray KBOs (Fig. 1). The Neptune Trojans do not appear to have the extreme red colors shown by objects in the low-inclination ("cold"), high-perihelion classical Kuiper Belt (*19*, *20*).

To determine whether the Neptune Trojans are drawn from the same distribution of colors as are other small-body populations, we performed some statistical tests. First, we note that about 30% of KBOs have colors similar to those of the Neptune Trojans (*21*). From this information, we find that there is less than a $(0.3)^4 \sim 1\%$ chance of observing the Neptune Trojan colors we found if they are drawn from the same color distribution as the KBOs. We further performed a Monte Carlo Kolmogorov-Smirnov (K-S) test on the colors (Table 2). We first determined the maximum cumulative distance (D-statistic) of the four Neptune Trojan colors when ranking them versus each population individually. A Monte Carlo simulation was used to estimate the significance of the D-statistic because of our small data set, for which common analytic approximations on the significance fail when the two sample populations are not of equal size. We randomly chose

Table 1. Trojan orbital elements and sizes. Orbital data shown for the known Neptune Trojans are the semimajor axis (*a*), inclination (*i*), and eccentricity (*e*). Also included are the median Jupiter Trojan data. All four known Neptune Trojans have been observed during at least two oppositions, including the Neptune Trojans discovered in 2005 from recovery observations in June 2006 using GMOS on the Gemini North 8.2-m telescope. The orbital data are from the Minor Planet Center (supporting online text). The radii (*r*) of the objects were calculated assuming an albedo of 0.05 (0.1) using the equation $r = [2.25 \times 10^{16} R^2\Delta^2/p_R\phi(0)]^{1/2} 10^{0.2(m_{\odot} - m_R)}$ where R is the heliocentric distance in AU, Δ is the geocentric distance in AU, $m_{\odot}$ is the apparent red magnitude of the sun (−27.1), p_R is the red geometric albedo, m_R is the apparent red magnitude of the Trojan, and $\phi(0) = 1$ is the phase function at opposition. Our Neptune Trojan discovery survey used the Magellan-Baade 6.5-m telescope with the 0.2 square degree wide-field of view Inamori Magellan Areal Camera and Spectrograph (IMACS) for imaging. Survey data were obtained on the nights of UT 16 and 17 October 2004 and 6, 7, and 8 October 2005. Conditions over the different nights varied, but all were photometric, although wind and humidity limited our efficiency. We covered about 12 square degrees of the sky within about 1 hour of the Neptune L4 point in right ascension and within 1.5° of the ecliptic and Neptune's orbit. The limiting magnitude of the images varied between about an R color band of 23.5 magnitudes in the poorest fields to near 25th magnitude in the best fields, with the average being between 24 and 24.5 magnitudes. We used a wide filter covering the typical V and R color bands for discovery.

Name	*a* (AU)	*i* (deg)	*e*	*r* (km)
2001 QR_{322}	30.14	1.3	0.03	70 (50)
2004 UP_{10}	30.08	1.4	0.03	50 (35)
2005 TN_{53}	30.05	25.1	0.07	40 (30)
2005 TO_{74}	30.05	5.3	0.06	50 (35)
Median Jup	~5.2	10.9	0.07	

Table 2. Optical photometry. The mean colors of individual Neptune Trojans and some dynamical classes with the K-S test confidence of difference probabilities (K-S Diff) from Neptune Trojans are shown: Neptune Troj, Neptune Trojans; Jupiter Troj, Jupiter Trojans; KBOs, Kuiper Belt objects; cold KBOs, classical KBOs with $i < 10°$ and KBOs with perihelion > 40 AU; Centaur blue, the blue lobe of the bimodal centaur distribution; Centaur red, the red lobe of the bimodal Centaur distribution; and Irr sats, Irregular satellites of the giant planets. Uncertainties listed for each dynamical class are the standard deviation showing the broadness of the color distribution. The standard deviation of the mean colors are much smaller, around 0.02 magnitudes. Color references are in the text.

Name	m_R (mag)	B-V (mag)	V-R (mag)	R-I (mag)	B-I (mag)
2001 QR_{322}	22.50 ± 0.01	0.80 ± 0.03	0.46 ± 0.02	0.36 ± 0.03	1.62 ± 0.04
2004 UP_{10}	23.28 ± 0.03	0.74 ± 0.05	0.42 ± 0.04	0.46 ± 0.05	1.63 ± 0.06
2005 TN_{53}	23.73 ± 0.04	0.82 ± 0.08	0.47 ± 0.07	0.47 ± 0.09	1.75 ± 0.10
2005 TO_{74}	23.21 ± 0.03	0.85 ± 0.06	0.49 ± 0.05	0.42 ± 0.06	1.76 ± 0.07
	K-S Diff				
Neptune Troj	0%	0.80 ± 0.05	0.46 ± 0.03	0.43 ± 0.05	1.69 ± 0.08
Jupiter Troj	45%	0.74 ± 0.05	0.45 ± 0.03	0.43 ± 0.03	1.63 ± 0.09
KBOs	99.2%	0.90 ± 0.15	0.57 ± 0.12	0.55 ± 0.13	2.08 ± 0.38
Cold KBOs	99.99%	0.98 ± 0.14	0.64 ± 0.10	0.57 ± 0.07	2.20 ± 0.29
Centaurs	98%	0.92 ± 0.21	0.59 ± 0.15	0.58 ± 0.12	2.08 ± 0.47
Centaur blue	15%	0.72 ± 0.05	0.45 ± 0.04	0.49 ± 0.07	1.66 ± 0.14
Centaur red	99.99%	1.11 ± 0.08	0.72 ± 0.05	0.67 ± 0.08	2.50 ± 0.20
Comets	95%	0.79 ± 0.05	0.49 ± 0.05	0.49 ± 0.12	1.78 ± 0.14
Irr sats	55%	0.78 ± 0.11	0.47 ± 0.09	0.39 ± 0.12	1.63 ± 0.20

four colors 100,000 times from each non–Neptune Trojan population color distribution and determined the D-statistic each time in the same manner as we did using the four Neptune Trojans. The confidence limits in Table 2 represent the percentage of times the randomly drawn colors' D-statistic was lower than the actual Neptune Trojan colors' D-statistic for each population.

To date, the known Neptune Trojan population is too scant to formally reject them as being from the same color distribution as the KBOs with only a 99.2% confidence. However, we find the Neptune Trojan colors are different from the extremely red classical low-inclination and distant perihelion KBOs at the 99.99% confidence level. We also find they are incompatible with the red lobe of the Centaur bimodal color distribution at the 99.99% confidence level (Fig. 1). The Neptune Trojan colors are most consistent with the blue lobe of the possible bimodal Centaur color distribution (*18*). Further, the Neptune Trojans are also consistent with the color distributions of the Jupiter Trojans (*22*), irregular satellites (*23*), and possibly the comets (*24*).

These colors are consistent with a similar origin for the Neptune Trojans, the Jupiter Trojans, irregular satellites, and dynamically excited gray Kuiper Belt population. These populations may have been subsequently dispersed, transported, and trapped in their current locations during or just after the planetary migration phase (*25*, *26*). The Neptune Trojans are too faint to efficiently observe spectroscopically with current technology. The slightly red surface color observed on some outer solar system objects can be reproduced with many different compounds. Most interpretations allow about two-thirds of the reflectance gradient to be attributed to Triton tholins and one-third to ice tholins, which can be produced by bombarding simple organic ice mixtures with ionizing radiation (*21*, *27*). However, other models have used no organic ices but simple Mg-rich pyroxene (*28*).

If we assume albedos for the known Trojans, we can estimate their effective radii from our photometric measurements (Table 1). The optical flux density of sunlight scattered from a solid object follows $f \propto D^2/R^4$, where D is the diameter and R is the heliocentric distance of the object. When assuming an albedo of 0.05 or 0.1 (for values in parentheses), respectively, we find the four known Neptune Trojans have radii from about 40 (30) to 70 (50) km. These sizes are comparable to the largest known Jupiter Trojans, which have albedos around 0.05.

We can estimate the expected number of L4 Neptune Trojans with sizes larger than our smallest discovered object, 2005 TN_{53}, by assuming that the Trojans are equally distributed throughout the L4 Trojan cloud with identical albedos. We discovered three Trojans in about 12 square degrees. If the Neptune Trojan cloud stretched 30° in right ascension and 50° in declination near the L4 Neptune point, it would cover 1500 square degrees. Using Poisson counting statistics, about 400^{+250}_{-200} Neptune Trojans with radii greater than about 40 (30) km are expected in the L4 Neptune cloud assuming albedos of 0.05 (0.1). The Jupiter Trojan population is complete to this size level. The Jupiter Trojan L4 cloud contains 30 (70) objects with radii greater than about 40 (30) km, whereas the L5 point harbors only about 20 (40) such objects. Depending on the albedos of the Neptune Trojans, we find they are between about 5 and 20 times more abundant at the large sizes than the Jupiter Trojans.

Through our Neptune Trojan survey and the findings discussed in (*10*), we find that the large Neptune Trojan size distribution may be quite steep (size distribution $q \sim 5.5$, where $n(r)dr \propto r^{-q}dr$ is the differential power-law radius distribution, with $n(r)dr$ the number of Neptune Trojans with radii in the range r to $r + dr$), which is very similar to the large Jupiter Trojans' size distribution. We found no faint Neptune Trojans (apparent red magnitude $(m_R) > 24$), even though our survey was sensitive to them. This may indicate a shallowing in the size distribution for the smaller Neptune Trojans, as also observed for the Jupiter Trojans (*9*) and KBOs (*29*). This may be because the largest objects have not been collisionally disrupted, whereas the smaller objects have been continually fragmented throughout the age of the Solar System.

All four known Neptune Trojans were found in the L4 region. It is likely that the Neptune L5 (trailing) region has a population of Trojans, as is true for Jupiter, but for the next several years, the line of sight through the Neptune L5 region passes near the plane of our galaxy, making observations very difficult because of stellar confusion.

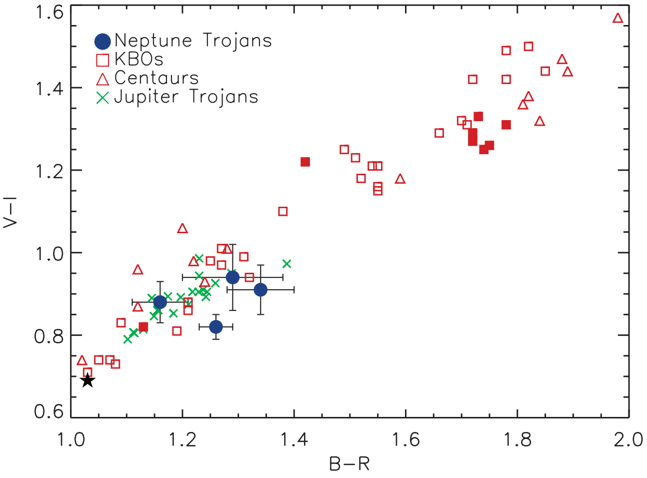


Fig. 1. Optical colors of the Neptune Trojans (solid blue circles) compared with the KBOs (red squares), Centaurs (red triangles), and Jupiter Trojans (green *x*'s). The extremely red KBOs and Centaurs are in the upper right; gray objects are in the lower left. The Neptune Trojans' colors are clearly distinct from the cold classical KBOs', which have inclinations less than 10° and KBOs with perihelions greater than 40 astronomical units (AU) (red solid squares). The Neptune Trojans are considerably more blue than the median KBO colors. The Jupiter Trojans and the blue lobe of the Centaur bimodal color distribution are similar to the Neptune Trojan colors. The color of the Sun is shown by the black star in the lower left corner. KBO and Centaur colors are from (*21*) and have typical error bars of 0.03 magnitudes. The Neptune Trojan color observations were obtained on photometric nights with the Magellan-Clay 6.5-m telescope with the LDSS-3 CCD camera using optical Sloan broad-band filters g′, r′, and i′. The Sloan colors were converted to the Johnson-Morgan-Cousins BVRI color system using transfer equations from (*30*). To verify the color transformation, we also observed an extremely red and a gray [(44594) 1999 OX3 and (19308) 1996 TO66, respectively] trans-Neptunian object. For each Trojan, we observed in all filters before repeating a filter. Each Neptune Trojan was observed on the nights of 2, 3, and 4 November 2005 universal time (UT). The standard star G158-100 was used for photometric calibration (*30*). Image quality was excellent with about 0.5 arc sec full width at half maximum seeing on each night. Exposure times were between 300 and 450 s for each image.

References and Notes

1. S. Peale, *Icarus* **106**, 308 (1993).
2. E. Chiang, Y. Lithwick, *Astrophys. J.* **628**, 520 (2005).
3. R. Gomes, *Astron. J.* **116**, 2590 (1998).
4. S. Kortenkamp, R. Malhotra, T. Michtchenko, *Icarus* **167**, 347 (2004).
5. F. Marzari, H. Scholl, *Icarus* **146**, 232 (2000).
6. A. Morbidelli, H. Levison, K. Tsiganis, R. Gomes, *Nature* **435**, 462 (2005).
7. F. Marzari, H. Scholl, *Icarus* **131**, 41 (1998).
8. H. Fleming, D. Hamilton, *Icarus* **148**, 479 (2000).
9. D. Jewitt, C. Trujillo, J. Luu, *Astron. J.* **120**, 1140 (2000).
10. E. Chiang *et al.*, *Astron. J.* **126**, 430 (2003).
11. J. Hahn, R. Malhotra, *Astron. J.* **130**, 2392 (2005).
12. D. Nesvorny, L. Dones, *Icarus* **160**, 271 (2002).
13. J. Horner, N. Evans, *Mon. Not. R. Astron. Soc.* **367**, L20 (2006).
14. J. Chambers, *Mon. Not. R. Astron. Soc.* **304**, 793 (1999).
15. J. Pollack *et al.*, *Icarus* **124**, 62 (1996).
16. E. Shoemaker, C. Shoemaker, R. Wolfe, in *Asteroids II*, R. Binzel, T. Gehrels, M. Matthews, Eds. (Univ. of Arizona Press, Tucson, AZ, 1989), pp. 487–523.
17. K. Tsiganis, R. Gomes, A. Morbidelli, H. Levison, *Nature* **435**, 459 (2005).
18. N. Peixinho *et al.*, *Icarus* **170**, 153 (2004).
19. S. Tegler, W. Romanishin, *Nature* **407**, 979 (2000).
20. C. Trujillo, M. Brown, *Astrophys. J.* **566**, L125 (2002).
21. M. Barucci, I. Belskaya, M. Fulchignoni, M. Birlan, *Astron. J.* **130**, 1291 (2005).

22. S. Fornasier *et al.*, *Icarus* **172**, 221 (2004).
23. T. Grav, M. Holman, B. Gladman, K. Aksnes, *Icarus* **166**, 33 (2003).
24. D. Jewitt, *Astron. J.* **123**, 1039 (2002).
25. R. Gomes, *Icarus* **161**, 404 (2003).
26. H. Levison, A. Morbidelli, *Nature* **426**, 419 (2003).
27. A. Doressoundiram *et al.*, *Astron. J.* **125**, 2721 (2003).
28. D. Cruikshank, C. Dalle Ore, *Earth Moon Planets* **92**, 315 (2003).
29. G. Bernstein *et al.*, *Astron. J.* **128**, 1364 (2004).
30. J. Smith *et al.*, *Astron. J.* **123**, 2121 (2002).

31. We thank D. Tholen and B. Marsden for orbital discussions; J.-R. Roy, I. Jorgensen, and the Gemini observatory staff for granting and executing directors discretionary observing time; as well as D. Jewitt and H. Hsieh for recovery observations. This paper includes data gathered with the 6.5-m Magellan telescopes located at Las Campanas Observatory, Chile, and is based in part on observations obtained at the Gemini Observatory, which is operated by the Association of Universities for Research in Astronomy, Inc., under a cooperative agreement with the NSF on behalf of the Gemini partnership: the National Science Foundation (United States), the Particle Physics and Astronomy Research Council (UK), the National Research Council (Canada), CONICYT (Chile), the Australian Research Council (Australia), CNPq (Brazil), and CONICET (Argentina). S.S.S. is supported by NASA through the Hubble Fellowship awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA. C.A.T. is supported by the Gemini Observatory, which is operated by the Association of Universities for Research in Astronomy, Inc., on behalf of the international Gemini partnership of Argentina, Australia, Brazil, Canada, Chile, the UK, and the USA.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1127173/DC1
Table S1

8 March 2006; accepted 6 June 2006
Published online 15 June 2006;
10.1126/science.1127173
Include this information when citing this paper.

Clonal Adaptive Radiation in a Constant Environment

Ram Maharjan, Shona Seeto, Lucinda Notley-McRobb, Thomas Ferenci*

The evolution of new combinations of bacterial properties contributes to biodiversity and the emergence of new diseases. We investigated the capacity for bacterial divergence with a chemostat culture of *Escherichia coli*. A clonal population radiated into more than five phenotypic clusters within 26 days, with multiple variations in global regulation, metabolic strategies, surface properties, and nutrient permeability pathways. Most isolates belonged to a single ecotype, and neither periodic selection events nor ecological competition for a single niche prevented an adaptive radiation with a single resource. The multidirectional exploration of fitness space is an underestimated ingredient to bacterial success even in unstructured environments.

Abundant bacterial variety occurs in most environments (*1*) and is believed to have arisen from adaptive radiation to the myriad of structured environmental and biotic niches and specialization on alternative resources (*2*, *3*). Thus, an environment with multiple niches results in more obvious diversification of experimental bacterial populations than a homogeneous environment (*4*, *5*). Yet the other inference from this ecological view of adaptation, that a uniform environment with a single resource leads to low sympatric diversity, has not been tested exhaustively. Understanding the rapidity and limits of bacterial diversification in defined environments would be of benefit in modeling everything from infection progression to the stability of large-scale industrial fermentations.

Although many mutations occur in large bacterial populations, a low phenotypic diversity is expected because of purges involving fitter mutants, called periodic selection events (*6*, *7*). Mutational periodic selection events correlate with purifying sweeps by strongly beneficial mutations (*8*). Still, it is questionable whether mutational selections maintain low diversity in clonal populations (*9*), and several long-term experimental evolution cultures resulted in stable coexistence of clones (*10*, *11*). The generation of new ecotypes and niches is one possible source of radiation in a restricted environment (*12*), with the evolution of cross-feeding polymorphisms as an example (*10*). The temporal and nutrient-concentration fluctuations in other intensively analyzed experimental evolution experiments (*11*) also cannot eliminate the possibility of specialization for unidentified niches. We explored the metabolic, phenotypic, genotypic, and ecotypic divergence in an evolving chemostat population with a constant unstructured environment and a single resource.

A screen for diversification was developed by using an *Escherichia coli* strain with a reporter gene that can detect several types of regulatory change, resulting in improved transport properties and large fitness benefits in glucose-limited chemostats [strain BW2952 (*13*–*15*)]. As described in (*16*), the uniformity of populations could be tested by assaying the *malG-lacZ* fusion activity, glycogen staining, and methyl α-glucoside (α−MG) sensitivity, which detects increased glucose uptake (*17*). Mutations affecting *rpoS*, *mlc*, *malT*, and *ptsG* influence one or several of the assayed traits to differing extents (*13*). Bacteria were grown at a dilution rate (*D*) of 0.1 hour^{-1} (a doubling time of 6.9 hours) with a population size of 1.6×10^{10} and sampled over 28 days. *rpoS* mutations, which confer a large fitness advantage (*18*), initially swept the population and led to a rapid elimination of parental *rpoS*$^+$ bacteria (Fig. 1). Several identified (*13*) and unidentified sweeps followed the *rpoS* sweep in the dominant *rpoS* subpopulation, but the uniformity of the sampled isolates stayed high in the first 2 weeks of culture. Subsequently, and especially after 17 days, the population diversified to reveal multiple combinations of properties (Fig. 1). Simultaneously, *rpoS*$^+$ bacteria again became a substantial proportion of the population. Similar diversity and recovery of *rpoS*$^+$ bacteria was found in three other replicate populations (respectively, 41%, 38%, 70%, and 35% *rpoS*$^+$) and prompted a more detailed analysis of late culture samples.

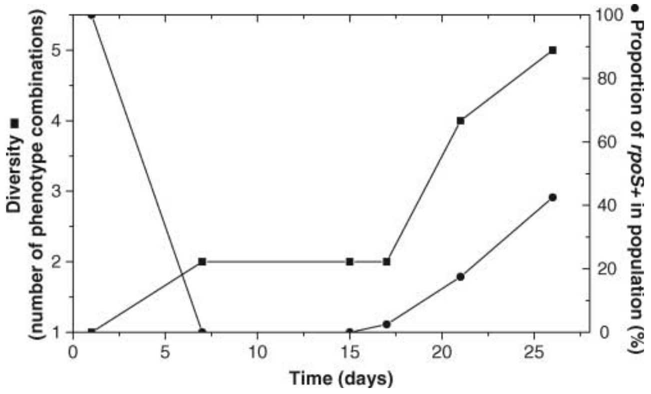


Fig. 1. Time course of changes in an evolving bacterial population. *E. coli* strain BW2952 was grown at $D = 0.1\ \text{hour}^{-1}$ in a glucose-limited chemostat as described in (*13*). Daily samples were analyzed for changes in *rpoS* (circles) (*16*, *34*). The diversity index had its basis in assays of α-MG sensitivity and *rpoS* status (both yes-no traits) and in the *malG-lacZ* activity shown in table S1 as the third trait. The fusion activity was divided into low (<200 units, wild type), intermediate (300–600 units), and high (700 units) ranges. The number of combinations of shared characters in 40 isolates tested is shown at each time point (squares).

School of Molecular and Microbial Biosciences, University of Sydney, Sydney, New South Wales 2006, Australia.

*To whom correspondence should be addressed. E-mail: tferenci@mail.usyd.edu.au

The isolates from day 26 were screened for eight further phenotypic and genotypic characteristics (table S1). All of these characteristics changed under prolonged glucose limitation. Growth yields on glucose were tested because of changes in other studies (*15*) as were outer membrane profiles (Fig. 2A) associated with altered porins and antibiotic susceptibilities (*19*). Aminotriazole (AT) sensitivity probed altered concentrations of ppGpp, an important alarmone of *E. coli* (*20*), because of possible effects of this global regulator on *rpoS*-related expression (*21*) and because of ppGpp-related *spoT* changes in other evolving populations (*22*). Variation in the aggregation of cultures was also observed (Fig. 2B) and compared. The characterizations included sequencing of the *rpoS* and *mgl* mutations (*13*) to differentiate alleles in the one culture. Table S1 summarizes 11 separate properties of the parental strain and 41 isolates.

The most prominent outcome of this analysis was the level of diversity among coevolved strains. Incorporating the data in Table S1 into a nearest neighbor–joining dendrogram revealed multiple branched clusters in the population (Fig. 3). One broad group included all *rpoS* mutants, with different *rpoS* alleles in the subclusters. The *rpoS*+ isolates differed in α−MG and AT sensitivity, so that population partitioned into three distinct approaches to global gene regulation associated with RpoS and ppGpp. The magnitude of the diversification in Fig. 3 suggests that a clonal bacterial population essentially became a collection of individual lineages in about 90 bulk generations under nutrient stress. The closest approximation to this level of diversification is the Lenski long-term *E. coli* lineages, where extensive insertion sequence rearrangements were observed over 10,000 generations (*23*) without recognition of the extensive phenotypic divergence described here.

The growth yields of the 41 isolates on glucose varied markedly (table S1). The distinct yields suggested at least four parallel approaches to glucose conversion into biomass and associated metabolic adaptations. The different yield classes produced different amounts of acetate, CO_2, and biomass from the equivalent amount of glucose; only one isolate produced acetate in a glucose-limited chemostat; and no other fermentation product was evident in pure cultures of other isolates (*16*). A cross-feeding polymorphism (*10*) was not common in this population within 90 generations.

Increased outer membrane permeability is an adaptive strategy for bacteria growing with low extracellular glucose concentrations (*24*). Analysis of outer membranes revealed five combinations of protein changes in individuals of the one population (table S1 and Fig. 2A). The changes were associated with increased membrane permeability, because there was an increased sensitivity to one or more antibiotics and detergent (table S1) (*19*). Individually inactivating the genes encoding outer membrane proteins (OmpF, OmpC, LamB, PhoE, and OmpG) in representative isolates indicated that the five groupings based on protein banding patterns underestimated the number of permeability-related changes in the population (table S2). Two class II isolates responded differently to *ompF* knockouts, indicating distinct mutational histories. Inactivation of each of the above proteins individually (table S2) did not reduce the elevated susceptibility to rifampicin. As with the regulatory alterations, there was a multiplicity of parallel permeability adaptations in a single population and even within a single cluster (compare BW3767 with BW4004).

The sedimentation of many isolates in a static culture (Fig. 2B) was also indicative of surface changes. Microscopic investigation indicated a tendency to aggregate in sedimenting bacteria (Fig. 2C). Nevertheless, in stirred chemostats, the isolates stayed in suspension as individual or dividing cells (Fig. 2C), so it is unlikely they formed a novel niche under the selection conditions. The clumping phenotype was not linked to any regulatory or growth-yield grouping (table S1) and so was not associated with a particular growth strategy. Surprisingly, 7 of 41 isolates remained in suspension better than the ancestor (Fig. 2D), indicative of at least four parallel surface or density changes. It remains to be established whether the sedimentation changes constitute a benefit under glucose limitation.

Did ecological specialization determine overall diversity? There were constant environmental conditions in the chemostat culture and no obvious fractionation of the population spatially or through adhesion. None of the 41 bacterial isolates showed increased adhesion to glass (*16*), and no obvious wall growth was observed in the selection chemostats. Temporal fluctuations were at best 20 to 25 s, between the drops added at the slow dilution rate used in the culture. There was no lysis (transient reductions in density) in the chemostats. We did not detect the production of secondary metabolites besides acetate in one isolate (*16*). In the absence of alternative niches, evidence for possible divergence through ecotypic changes was sought from competition experiments between members of the population (*4*, *7*).

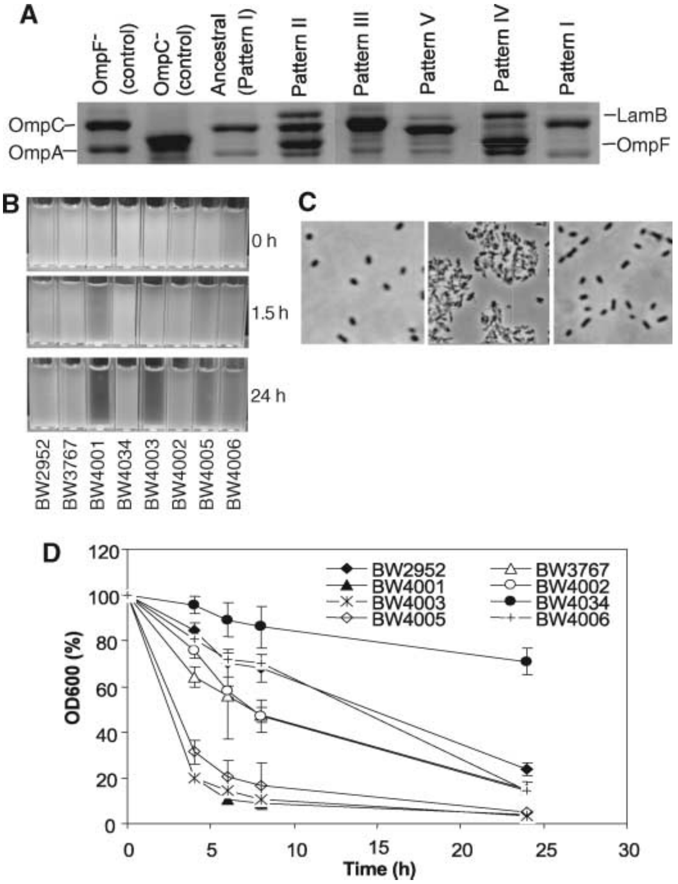


Fig. 2. Phenotypic changes in evolved isolates after 26 days of selection. (**A**) The outer membrane proteins of 41 evolved clones of *E. coli* under glucose limitation were analyzed by using the SDS-urea electrophoresis method described in (*35*). The positions of porin protein bands OmpF and OmpC were identified by comparing outer membrane protein (OMP) profiles of mutants lacking either OmpF or OmpC. One example from each of the five groups of OMP patterns observed in the 41 isolates (table S1) is shown. The OmpF⁻ control was strain MH513 and the OmpC⁻ control was MH225. The ancestral strain produces pattern I, and chemostat-evolved isolates BW3767, pattern II; BW4001, pattern III; BW4011, pattern IV; BW4027, pattern V; and BW4003, pattern I. (**B**) The sedimentation of cultures was initially observed as shown in the cuvettes. (**C**) The rapidly sedimenting BW4001 culture was observed by phase-contrast microscopy (Olympus BH, Olympus Optical Company, Tokyo, Japan) directly from a chemostat culture (left), as well as after 6 hours of standing (center). (Right) The parental BW2952 strain after 6 hours of standing. (**D**) The different rates of sedimentation of representative isolates from table S1. Error bars indicate standard deviations based on three replicants.

Isolates from various clusters in Fig. 3 were tested in direct competition experiments with ancestral bacteria. The marker used to resolve competing populations did not change the frequency of ancestor (Fig. 4A) (BW2952). Of 10 isolates tested, 8 displaced the parental strain from a glucose-limited chemostat, indicating fitness in the original selection niche. The displacement by BW3767 was not due to cross-inhibition of ancestor, because there was no difference in the growth of either strain in spent medium from a BW3767 culture. Elimination is compatible with the starvation of ancestor for glucose and washing out in the presence of fitter isolates.

Interestingly, BW4005 was less fit than the ancestor and likely to belong to a new ecotype. Presumably BW4005 used resources in the long-term culture absent from the reconstituted chemostat. BW4005 and BW4003 were in the same phenotypic cluster in Fig. 3 but had different fitness properties, indicating yet unidentified divergence(s) between the strains. Because BW4003 did not fully eliminate the ancestor, it was possible that this strain also used an alternative resource. To further test whether BW4003 and isolates from other phenotypic groupings showed fitness properties consistent with ecotype divergence, we prepared competitions between evolved strains from the different clusters in Fig. 3 (Fig. 4B). The selection coefficients in Fig. 4B are shown for pairwise competitions between four isolates when the isolates start at 1% and 99% abundances, respectively. No two strains had identical fitness properties, consistent with ongoing population shifts evident from Fig. 1. BW4003 was the least fit against all strains, consistent with the weaker competition against ancestor (Fig. 4A). Only BW4001 exhibited frequency-dependent fitness properties in competition with each of the other strains. Applying the definition of an ecotype (*7*), that an adaptive mutant from within an ecotype outcompetes to extinction other strains of the same ecotype, we find that three of the four divergent strains in Fig. 4B are competing for a single niche. We have not identified alternative niches, but bacteria such as BW4005 and BW4001 may be evolving toward alternative ecotypes. Also, the relatively small sample in Fig. 4B may not have uncovered other variations within the large clusters in Fig. 3.

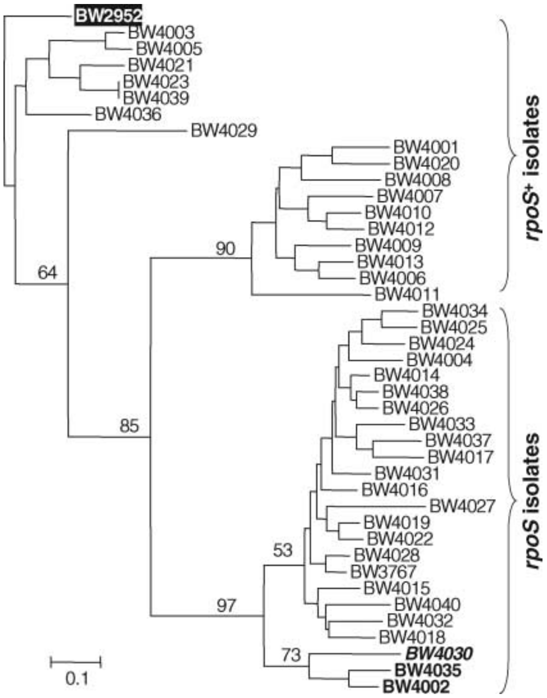


Fig. 3. Phylogenetic relationships among *E. coli* isolates evolved in a single population. Sympatric divergence of evolved clones was analyzed by the neighbor-joining method rooted to the ancestral strain, BW2952 (black box and white type), as described in (*36*). The dendrogram has its basis in the 11 characteristics of the 41 isolates shown in table S1. Isolates printed in bold or italic type within the large *rpoS* cluster indicate isolates with distinct *rpoS* alleles (see table S1 for details). Numbers on the branches indicate bootstrap values based on 1000 replications.

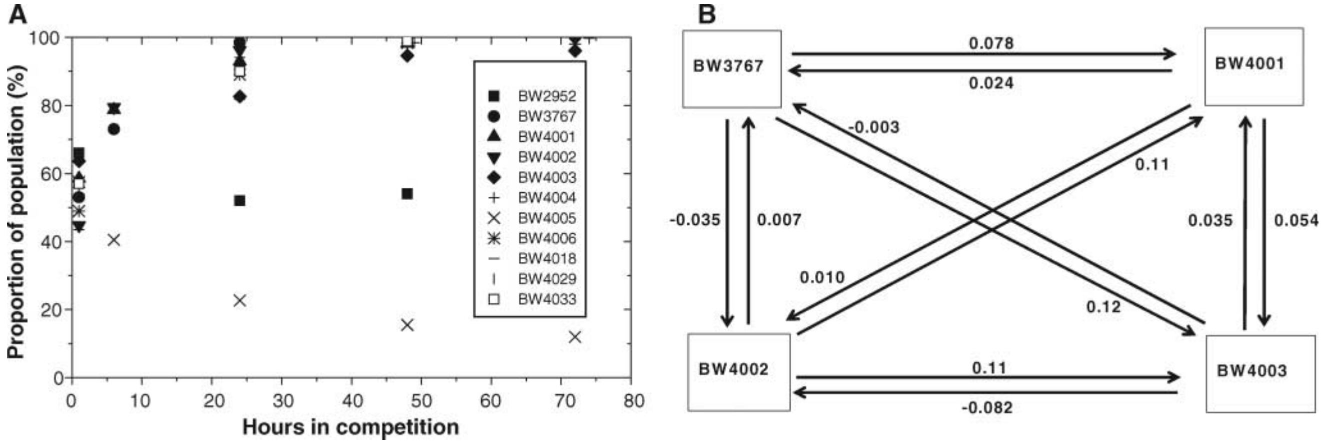


Fig. 4. Competitive interactions between different isolates under glucose limitation. (**A**) The population composition was followed when two strains, grown independently for 24 hours under glucose limitation (D = 0.1 hour^{-1}), were mixed in equal proportions and the culture maintained under glucose limitation at the same dilution rate. The strains were differentiated by using T5 resistance as a neutral marker in the ancestral strain. T5-sensitive BW2952 and its T5-resistant derivative (BW3494) were equally fit during these experiments (solid squares). The composition of the cultures when BW3494 was mixed with strains from the various clusters in Fig. 3 is shown. Each experiment was repeated two to three times with similar patterns obtained. (**B**) Each of the isolates was competed pairwise against each of the other three strains. The inocula and mixing were performed as in (A). Duplicate competitions were initiated with chemostat-adapted bacteria mixed at 1% and 99% ratios of each strain. The arrows point away from the 1% or low-abundance start point for each isolate. The reported s values, or selection coefficients (*14*), are shown for each direction for each pair, with positive values showing increasing proportion or higher fitness in the competition experiments. Negative values indicate decreasing proportions in the competitions. The population estimates, as described in (*16*), have their basis in the means of the strain counts, including those obtained with reciprocal markers; the variation found was <7.5% standard deviation.

Overall, the 41 isolates contained at least three regulatory, four metabolic, four aggregation-related, and five different membrane permeability solutions. The observed combinations must reflect the multiplicity of physiological choices available to bacteria adapting to a glucose limitation. Evidently, an adaptive radiation can take place in a near-constant environment, and bacterial populations do not fully subscribe to the competitive exclusion principle (*2*). A rare, highly beneficial mutational advantage may still arise in long-term chemostat populations and purge diversity, but analysis of four parallel populations did not exhibit such sweeps beyond the first 2 weeks of culture.

How do we explain the number of mutations in the 26-day population? Judging from the sequence changes in table S1 and that the MG and AT sensitivities and different outer membrane changes are due to distinct mutations, most isolates contain at least four or five beneficial as well as an unknown number of neutral mutations. The multiplicity of mutations collected within 26 days does not agree with the Drake estimate of long-term bacterial mutation rates [0.0033 mutations per division (*25*)], which would yield only 0.3 mutations per genome in 90 generations. The mutations occurred in the absence of detected mutator cells in this culture (*16*), but two factors contributed to high mutation supply. First, the mutation frequency is about 30-fold elevated in a glucose-limited chemostat above that in nutrient-excess bacteria (*13*). Secondly, the successful bacteria will have undertaken more than 90 divisions, because the spread of each beneficial mutation must lead to transiently faster growth rates within the chemostat population (Fig. 4A).

There is extensive individuality of the isolates by day 26 (Fig. 3). Yet the initial sweeps did not introduce phenotypic diversity into the population (Fig. 1), although several alleles of mutated genes [as in *rpoS*, *mgl*: (table S1)] were present. Consistent with longer-term studies of bacterial adaptation, the major gains in fitness, such as the *rpoS* sweep, occurred early in the life of a population (*26*). Clonal interference, or competition between clones with different beneficial mutations in the population (*27*), may have slowed the emergence of multiple types in the first weeks. Beyond the early gains, incremental changes to properties such as membrane permeability improved fitness, but probably in small steps. The recovery of *rpoS*$^+$ bacteria did not occur as a rapid sweep, and the mutations accumulating in the two lineages of *rpoS*$^+$ bacteria were not beneficial enough to displace coevolving genotypes. Purifying periodic selections are inherently unlikely when weakly beneficial mutations arise in large populations subject to persistent selection (*7*, *9*). To explain the increasing diversity, we propose that mutations of small fitness benefit swamp clonal interference or purifying periodic selection events. With isolates such as BW4001, negative frequency-dependent selection also operates to maintain diversity in the chemostat population, as discussed elsewhere (*28*).

The major radiation apparent in a single clone may seem inconsistent with the observation that bacteria in nature can be classified into groupings discernable as species. It needs to be remembered that our study excluded the acquisition of foreign genes. Lateral gene transfer (*29*) probably does superimpose a purging effect and maintenance of some order in bacterial relationships, because such transfers are rarer than the mutational sources of genetic change studied above. Periodic selection events in nature may indeed mostly depend on lateral gene transfer. It is also true that the adaptive radiation in the chemostat involved evolutionary overspecialization, with its strong tradeoff costs. Many of the adaptations involved antagonistic pleiotropy. For example, the increased detergent sensitivity of many isolates reflects increased membrane permeability beneficial specifically under chemostat conditions (*18*). These outer membrane changes would result in killing by bile salts in the normal habitat of *E. coli.* The RpoS mutations and reduced ppGpp would also cause problems in the transition to more stressful environments (*17*). Nevertheless, the large pool of genetic variation available through clonal divergence may be the source material for rarer lateral transfer in times of stress. Much of the microevolution seen within bacterial species (*30*) may be ultimately sourced to clonal diversification.

The speed and scale of the observed radiation has implications for evaluating bacterial success in all situations. Multidirectional divergence is relevant to bacteria in populations during persistent infections (*31*), facing a new environment, or crossing to a new host. In this context, it is relevant that mutation supply is limiting in pathogen populations, which are not as large as the population we investigated, and that mutations in mutator genes are often associated with pathogenesis (*32*). Our results suggest ecological specialization for multiple niches is not essential for bacterial diversity (*3*, *33*) and that mutational periodic selections are unlikely to ensure the purity of bacterial species in the absence of lateral gene transfer. Lastly, our results suggest that sharing of a niche by a large number of diversifying members of the same species is a feasible evolutionary strategy. A single fitness solution, or survival of the fittest, is not the only answer in a competitive environment.

References and Notes

1. J. C. Venter *et al.*, *Science* **304**, 66 (2004); published online 4 March 2004 (10.1126/science.1093857).
2. G. Hardin, *Science* **131**, 1292 (1960).
3. R. Kassen, P. B. Rainey, *Annu. Rev. Microbiol.* **58**, 207 (2004).
4. P. B. Rainey, M. Travisano, *Nature* **394**, 69 (1998).
5. B. Kerr, M. A. Riley, M. W. Feldman, B. J. M. Bohannan, *Nature* **418**, 171 (2002).
6. K. C. Atwood, L. K. Schneider, F. J. Ryan, *Cold Spring Harb. Symp. Quant. Biol.* **16**, 345 (1951).
7. F. M. Cohan, in *Microbial Genomes*, C. M. Fraser, T. D. Read, K. E. Nelson, Eds. (Humana, Totowa, NJ, 2004), pp. 175–194.
8. L. Notley-McRobb, T. Ferenci, *Genetics* **156**, 1493 (2000).
9. R. Korona, *Genetics* **143**, 637 (1996).
10. D. S. Treves, S. Manning, J. Adams, *Mol. Biol. Evol.* **15**, 789 (1998).
11. D. E. Rozen, R. E. Lenski, *Am. Nat.* **155**, 24 (2000).
12. F. M. Cohan, *Annu. Rev. Microbiol.* **56**, 457 (2002).
13. L. Notley-McRobb, S. Seeto, T. Ferenci, *Proc. R. Soc. London Ser. B* **270**, 843 (2003).
14. D. Dykhuizen, D. Hartl, *Evolution Int. J. Org. Evolution* **35**, 581 (1981).
15. R. B. Helling, C. N. Vargas, J. Adams, *Genetics* **116**, 349 (1987).
16. Materials and Methods are available on *Science* Online.
17. K. Manché, L. Notley-McRobb, T. Ferenci, *Genetics* **153**, 5 (1999).
18. T. King, A. Ishihama, A. Kori, T. Ferenci, *J. Bacteriol.* **186**, 5614 (2004).
19. E. Zhang, T. Ferenci, *FEMS Microbiol. Lett.* **176**, 395 (1999).
20. K. E. Rudd, B. R. Bochner, M. Cashel, J. R. Roth, *J. Bacteriol.* **163**, 534 (1985).
21. K. Kvint, A. Farewell, T. Nystrom, *J. Biol. Chem.* **275**, 14795 (2000).
22. T. F. Cooper, D. E. Rozen, R. E. Lenski, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1072 (2003).
23. D. Papadopoulos *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3807 (1999).
24. T. Ferenci, *FEMS Microbiol. Rev.* **18**, 301 (1996).
25. J. W. Drake, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7160 (1991).
26. S. F. Elena, R. E. Lenski, *Nat. Rev. Genet.* **4**, 457 (2003).
27. D. E. Rozen, J. de Visser, P. J. Gerrish, *Curr. Biol.* **12**, 1040 (2002).
28. M. L. Friesen, G. Saxer, M. Travisano, M. Doebeli, *Evolution Int. J. Org. Evolution* **58**, 245 (2004).
29. H. Ochman, J. G. Lawrence, E. A. Groisman, *Nature* **405**, 299 (2000).
30. E. J. Feil, *Nat. Rev. Microbiol.* **2**, 483 (2004).
31. J. V. Solnick, L. M. Hansen, N. R. Salama, J. K. Boonjakuakul, M. Syvanen, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2106 (2004).
32. J. E. LeClerc, B. Li, W. L. Payne, T. A. Cebula, *Science* **274**, 1208 (1996).
33. M. Feldgarden, D. M. Stoebel, D. Brisson, D. E. Dykhuizen, *Ecology* **84**, 1679 (2003).
34. L. Notley-McRobb, T. King, T. Ferenci, *J. Bacteriol.* **184**, 806 (2002).
35. X. Liu, T. Ferenci, *J. Bacteriol.* **180**, 3917 (1998).
36. G. M. Pupo, D. K. R. Karaolis, R. T. Lan, P. R. Reeves, *Infect. Immun.* **65**, 2685 (1997).

37. We thank R. Lan for advice on trees; G. Glaser for suggesting the AT assays; O. Tenaillon, D. Schneider, A. Holmes, and an anonymous reviewer for constructive comments on the manuscript; and the Australian Research Council for funding support.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1129865/DC1
Materials and Methods
SOM Text
Table S1 to S3
References

11 May 2006; accepted 7 June 2006
Published online 6 July 2006;
10.1126/science.1129865
Include this information when citing this paper.

Transcript-Assisted Transcriptional Proofreading

Nikolay Zenkin,[1]* Yulia Yuzenkova,[1] Konstantin Severinov[1,2,3]*

Fidelity of template-dependent nucleic acid synthesis is the main determinant of stable heredity and error-free gene expression. The mechanism (or mechanisms) ensuring fidelity of transcription by DNA-dependent RNA polymerases (RNAPs) is not fully understood. Here, we show that the 3′ end–proximal nucleotide of the nascent transcript stimulates hydrolysis of the penultimate phosphodiester bond by providing active groups and coordination bonds to the RNAP active center. This stimulation is much higher in the case of misincorporated nucleotide. We show that during transcription elongation, the hydrolytic reaction stimulated by misincorporated nucleotides proofreads most of the misincorporation events and thus serves as an intrinsic mechanism of transcription fidelity.

The mechanism of transcription is highly conserved in all living organisms. In the RNAP elongation complex, the 3′ end of the nascent RNA can occupy a post-translocated, a pretranslocated, or a backtracked state (Fig. 1A). In each of these states, the RNAP active center performs different reactions, i.e., the forward reaction of nucleoside triphosphate (NTP) addition or hydrolytic cleavage of the nascent RNA (Fig. 1A). Catalysis by the RNAP active center depends on two Mg^{2+} ions (*1*–*7*) that activate reacting groups and stabilize leaving groups during nucleophilic attack on the phosphorus (*4*, *8*). In cellular RNAPs, only one Mg^{2+} ion (MgI) is bound tightly in the active center. The other Mg^{2+} ion, MgII (*2*, *4*–*7*), is bound weakly, but its binding is stabilized by the triphosphate moiety of the incoming NTP (*5*). The hydrolytic transcript cleavage reaction, characteristic of pretranslocated and backtracked elongation complexes (*9*) (Fig. 1A), is slow compared to the forward RNA polymerization reaction, presumably because of poor binding of MgII (*4*).

Although the rate of misincorporation of nucleotides by RNAP is much slower than the rate of incorporation of correct nucleoside 5′-monophosphates (NMPs) (*10*, *11*), the relatively low selectivity of RNAP (*12*) makes misincorporation unavoidable, suggesting the existence of a proofreading mechanism. To define such a mechanism, 12 complexes with mismatched NMP at RNA 3′ end (misincorporated elongation complexes, MECs) modeling all possible misincorporation events and 4 correct complexes (correct elongation complexes, CECs) were assembled by means of 4 DNA templates that differed from each other only by a base pair at the +1 register (corresponding to the transcript 3′ end) and 4 5′ end–labeled RNAs that were identical except for the 3′-terminal base (*13*) (fig. S1). Complexes were assembled in the absence of Mg^{2+} and were therefore inactive.

Upon addition of Mg^{2+} to MECs, RNAP efficiently cleaved the penultimate (P2) phosphodiester bond. No P1 (ultimate phosphodiester bond) (Fig. 1A) cleavage was observed (Fig. 1, B and C), suggesting that MECs are backtracked by 1 base pair relative to the pretranslocated state (Fig. 1A). Cleavage of P2 was much slower in CECs than in the corresponding MECs (Table 1), as expected for active, not backtracked, complexes. However, because no P1 cleavage was observed in CECs (Fig. 1, B and C), stabilization of the backtracked state in MECs cannot explain preferential P2 cleavage, which was also observed with eukaryal and archeal RNAPs (*10*, *14*–*17*) and appears to be a general phenomenon.

Noncomplementary NTPs bind in the RNAP E-site close to the active center and stimulate P1 cleavage (*4*). Addition of nonhydrolyzable [to prevent any possibility of (mis)incorporation in the nascent RNA] NTP analog APcPP {adenosine-5′-[(α,β)-methyleno]triphosphate} (Fig. 1C), noncomplementary to the DNA template guanosine in register +2, led to stimulation of P1 cleavage in CECs. No such stimulation was observed in MECs, indicating that base-pairing of the RNA's 3′ end is required for NTP-assisted cleavage. Noncomplementary NTPs also inhibited P2

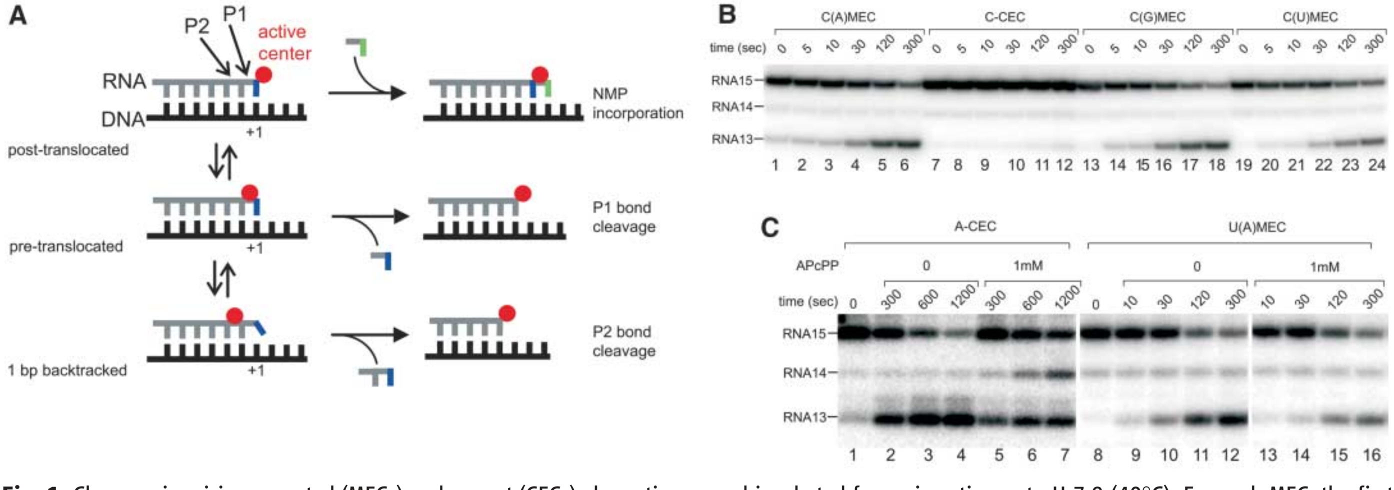


Fig. 1. Cleavage in misincorporated (MECs) and correct (CECs) elongation complexes. (**A**) Schematic representation of catalytic reactions characteristic of transcription elongation complexes in different states. The red circle represents the active center that contains two Mg^{2+} ions. (**B**) MECs (lanes 1 to 6, 13 to 24) and a corresponding CEC (lanes 7 to 12) (with CMP at the RNA 3′ end as an example) were supplied with 10 mM Mg^{2+} and incubated for various times at pH 7.9 (40°C). For each MEC, the first letter indicates misincorporated 3′ NMP, and the correct nucleotide that it replaces is indicated in parentheses (fig. S1). (**C**) CECs and MECs [A-CEC and U(A)MEC are shown as examples] were supplied with 15 mM Mg^{2+} and incubated for various times with or without 1 mM noncomplementary nonhydrolyzable NTP (APcPP).

[1]Waksman Institute, [2]Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, NJ 08854, USA. [3]Institute of Molecular Genetics, Russian Academy of Sciences, Moscow, 123182 Russia.

*To whom correspondence should be addressed at Waksman Institute, 190 Frelinghuysen Road, Piscataway, NJ 08854, USA. E-mail: nicserzen@mail.ru (N.Z.); severik@waksman.rutgers.edu (K.S.)

cleavage in both CECs and MECs in a dose-dependent manner (Fig. 1C). Noncomplementary NTPs are known to bind in the so-called E-site of the RNAP active center, the same site where initial interaction of correct NTPs with RNAP occurs. To explain the inhibitory effect, we postulate that in backtracked complexes, the 3′-terminal NMP also occupies the E-site (or an overlapping site, Fig. 2) and activates P2 cleavage in a way that is similar to P1 cleavage activation by noncomplementary NTP. Binding of NTP in the E-site displaces the transcript's 3′ end and destabilizes the backtracked state, thus inhibiting P2 cleavage.

Noncomplementary NTP activates P1 cleavage by stabilizing MgII through interaction with the β and γ phosphates (*4*, *5*). Because these phosphates are absent in the 3′-terminal NMP, MgII coordination and/or P2 cleavage may be stimulated by the terminal NMP itself. This hypothesis predicts that Mg^{2+} dependence of P2 cleavage, which reflects the complex affinity for MgII (*4*), should have a lower dissociation constant (K_d) than the intrinsic (unassisted by noncomplementary NTP) K_d of P1 hydrolysis (>100 mM) (*4*). This expectation was fulfilled for both MECs and CECs (Table 1). The lowest apparent K_d observed (8 mM) is close to the K_d of P1 hydrolysis stimulated by noncomplementary NTP (*4*). Thus, the 3′-terminal NMP, either matched or mismatched, increases the P2 cleavage velocity by increasing affinity for MgII.

A high rate of P2 cleavage could not be solely due to a decreased K_d for MgII, because cleavage velocity at saturating Mg^{2+} concentrations (k_{cat}) differed depending on the nature of 3′-terminal NMP (Table 1). For example, comparisons of complexes containing A, G, or U instead of correct C at the 3′ end (Table 1 rows 2, 7, and 15, correspondingly) reveal that the cleavage reaction k_{cat} values in different complexes differ significantly (0.14, 0.028, and 0.015 s^{-1}, respectively). This suggests that some groups of the transcript's 3′-end NMPs participate, directly or indirectly, in cleavage. Whereas the k_{cat} of P2 cleavage in MECs is determined by the properties of the reaction itself, in CECs it is strongly influenced by base-pairing of the 3′ end with the template strand, which affects the probability of backtracked state occupancy. To avoid this complication, we focused on MECs only (supporting online text).

High pH deprotonates the active water molecule stimulating phosphodiester hydrolysis by the RNAP active center. The stimulation depends on the reaction mechanism and should plateau at a pH equal to the system p*K* value. Therefore, if mismatched nucleotides were involved in cleavage, different profiles of cleavage reaction dependence on pH are expected for different MECs. This expectation was fulfilled (fig. S2). The shapes of pH curves were different from that of the previously reported P1 cleavage curve (*4*) (dotted line in fig. S2), indicating that the mechanism of P2 cleavage was distinct from intrinsic RNAP-catalyzed P1 hydrolysis. Whereas most P2 cleavage profiles plateaued at about pH 9.5, some had a different [U(C)MECs] plateau or even double (A-MECs) plateaus. Thus, different acid/base systems provided by the transcript 3′-terminal nucleotide participate in P2 cleavage in different MECs. The dependence of cleavage reaction properties for complexes containing misincorporated cytidine 5′-monophosphate (CMP) and uridine 5′-monophosphate (UMP) on the +1 DNA template-strand base may be explained by effects of local sequence-dependent deviations of nucleic acids structure near the active center on the reaction pathway (supporting online text).

To check which chemical groups of 3′-terminal NMP participate in MgII stabilization and P2 hydrolysis, we determined the cleavage reaction K_d and k_{cat} in MECs with RNAs containing chemical modifications in the phosphate, sugar, and base of the 3′-terminal nucleotide (Fig. 2). The results (Table 1 and table S1), discussed in detail in the supporting online text, are summarized below (see also fig. S3).

With misincorporated adenosine 5′-monophosphate, one of the P1 oxygens interacts with the 3′-hydroxyl, which in turn coordinates MgII. Another P1 oxygen orients the active water molecule. The 2′-hydroxyl does not participate in the reaction. N-7 of the purine ring coordinates MgII; the amino group in position 6 participates in water-molecule orientation or, alternatively, acts, together with nitrogen in position 1, as a general acid-base system. For misincorporated guanosine 5′-monophosphate (GMP), one of the P1 oxygens orients active water. The 2′ and 3′ hydroxyls do not participate in the reaction. N-7 of the base coordinates MgII. The amino group in position 2 fixes the GMP moiety, probably through interactions with the protein, making the reaction insensitive to local variations in nucleic acid structure.

Table 1. K_d for Mg^{2+} and k_{cat} of P2 cleavage for all possible CECs and MECs. All experiments were carried out in pH 7.9 (40°C). K_d and k_{cat} values were calculated with the Michaelis-Menten equation.

Complex	3′-end NMP of the RNA	Incorporated instead of	K_d (Mg^{2+}) (mM)	k_{cat} (s^{-1})
CEC	A	A	10	0.004
MEC		C	9	0.14
		G	8	0.12
		U	9	0.11
CEC	G	G	9	0.001
MEC		A	11	0.024
		C	15	0.028
		U	14	0.027
CEC	C	C	57	0.001
MEC		A	49	0.026
		G	15	0.029
		U	46	0.026
CEC	U	U	37	0.001
MEC		A	23	0.054
		C	8	0.015
		G	30	0.043

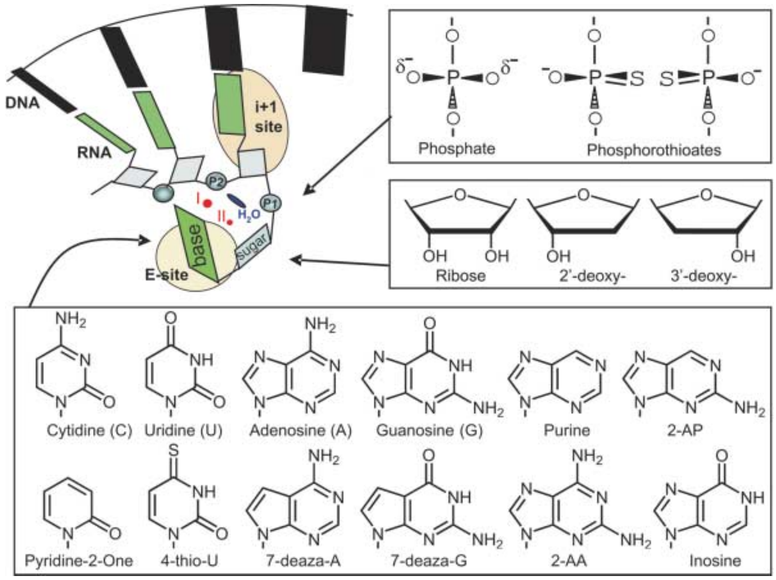


Fig. 2. Modifications of misincorporated 3′-terminal nucleotides used in this study. A schematic representation of the active center of RNAP in MEC that is consistent with our findings is shown on the left. Structures of modified bases, phosphate groups, and sugars are shown.

With misincorporated CMP, sequence dependence of the cleavage reaction in C-MECs is due to differences in the P1 bond orientation, which appears to be sensitive to local variations of nucleic acids structure. In one type of complex, P1 interacts with the 3′-hydroxyl, which coordinates MgII. In other complexes, this interaction is absent, and the 3′-hydroxyl does not chelate MgII. P1 also participates in a network of hydrogen bonding that positions the active water molecule. The 2′-hydroxyl is dispensable. Nitrogen in position 3 of the base chelates MgII. Finally, with misincorporated UMP, P1 interacts with the 3′-hydroxyl, positioning it to coordinate MgII or to orient the active water molecule. P1 also participates in coordination of the active water molecule. The 2′-hydroxyl is dispensable. The keto group in position 4 of the base either positions the water molecule or acts in concert with N-3 as a general base/acid.

Taken together, the results indicate that nucleotides that are misincorporated at the transcript 3′ end participate in their own excision. In contrast to the previously described stimulation of transcript cleavage by noncomplementary NTP (*4*), which can be regarded as "substrate-assisted catalysis" (*18*, *19*), the reaction described here represents "product-assisted catalysis" and, therefore, can directly affect transcription fidelity.

To show that excision of misincorporated NMP via P2 cleavage can prevent transcription past misincorporated NMP, we supplied MECs with NTP specified by the +2 register of the template and monitored transcript extension (Fig. 3). As noted for RNAPs from eukaryotes and archaea (*10*, *17*, *20*, *21*), the rate of incorporation of NTPs by MECs was much lower than by CECs (*7*, *22*) and was comparable ($k_{obs} \approx 0.03\ s^{-1}$ in the presence of 1 mM NTP) to the rate of P2 cleavage (Table 1). Presumably, slow elongation of misincorporated transcripts is due to stabilization of MECs in a backtracked state and to the occupancy of the primary NTP binding site, the E-site, by misincorporated NMP.

At 100 μM NTP, only 5 to 13% of MECs (30% for G-MEC) extended the RNA, whereas the rest of RNA was cleaved and, therefore, the misincorporated NMP was removed (Fig. 3B). In the presence of 1 mM NTP (a physiological concentration), ~30% of complexes (50% for G-MEC) extended past incorrect NMP, whereas the remainder underwent cleavage (Fig. 3B). When NTP was added together with transcript cleavage factor GreA, very low (except 20% for G-MEC) incorporation was detected, and mismatched NMP was removed (Fig. 3B). Thus, cleavage stimulated by misincorporated nucleotides is sufficient to proofread most misincorporation events. This activity is stimulated by transcript cleavage factors that were previously suggested to contribute to transcriptional fidelity (*10*, *12*, *17*, *21*) and that act by direct stabilization of MgII in the RNAP active site (*23*).

The importance of transcriptional proofreading for error-free gene expression was suggested (*24*). In addition, complexes containing misincorporated nucleotides elongate RNA slowly, which should impede expression of actively transcribed genes and may interfere with DNA replication. Cleavage factors cannot be solely responsible for removal of misincorporated nucleotides, because they are not essential for cells. Our results reveal a proofreading mechanism that may be sufficient to control transcription misincorporation in the absence of cleavage factors. The mechanism, which is likely evolutionarily conserved, also allows the removal of 2′-deoxy NMPs erroneously incorporated in RNA, because ribo and 2′-deoxy NMPs cleaved out with the same efficiency.

In the RNA-protein world, when RNAP was likely replicating RNA genomes (*25*), the relatively low fidelity of RNAP-catalyzed synthesis could not have been sufficient for stable maintenance of large RNA genomes in the absence of cleavage factors (*24*). A proofreading and repair mechanism similar to the one described here could have allowed a large RNA genome of the last common universal ancestor to exist.

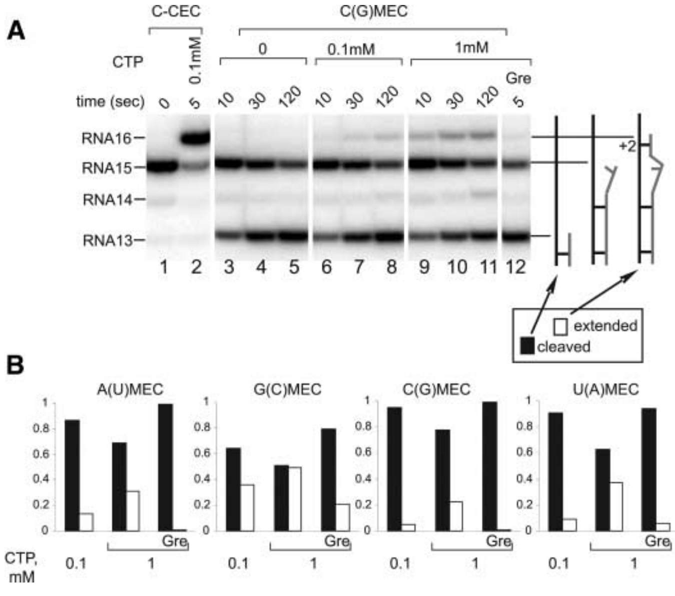


Fig. 3. P2 cleavage and transcriptional proofreading. (**A**) CECs and MECs [C-CEC (lanes 1 and 2) and C(G)MEC (lanes 3 to 12) are shown as examples] were supplied with 15 mM Mg^{2+} at pH 7.9 (40°C) and incubated for various times with or without different concentrations of CTP, specified by the +2 register of template DNA and 1 μM GreA. (**B**) Plots represent the relative amounts of MECs [A(U)MEC, C(G)MEC, G(C)MEC, and U(A)MEC are shown as examples] that incorporated NMP specified by the +2 register (white bars) and those that underwent P2 cleavage (black bars) in an experiment similar to that shown in (A).

References and Notes

1. T. A. Steitz, *Nature* **391**, 231 (1998).
2. P. Cramer, D. A. Bushnell, R. D. Kornberg, *Science* **292**, 1863 (2001).
3. D. G. Vassylyev *et al.*, *Nature* **417**, 712 (2002).
4. V. Sosunov *et al.*, *EMBO J.* **22**, 2234 (2003).
5. K. D. Westover, D. A. Bushnell, R. D. Kornberg, *Cell* **119**, 481 (2004).
6. H. Kettenberger, K. J. Armache, P. Cramer, *Mol. Cell* **16**, 955 (2004).
7. D. Temiakov *et al.*, *Mol. Cell* **19**, 655 (2005).
8. T. A. Steitz, J. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6498 (1993).
9. M. Orlova, J. Newlands, A. Das, A. Goldfarb, S. Borukhov, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4596 (1995).
10. M. J. Thomas, A. A. Platas, D. K. Hawley, *Cell* **93**, 627 (1998).
11. G. Bar-Nahum *et al.*, *Cell* **120**, 183 (2005).
12. D. A. Erie, O. Hajiseyedjavadi, M. C. Young, P. H. von Hippel, *Science* **262**, 867 (1993).
13. Materials and methods are available as supporting material on *Science* Online.
14. H. Guo, D. H. Price, *J. Biol. Chem.* **268**, 18762 (1993).
15. M. G. Izban, D. S. Luse, *J. Biol. Chem.* **268**, 12864 (1993).
16. S. K. Whitehall, C. Bardeleben, G. A. Kassavetis, *J. Biol. Chem.* **269**, 2299 (1994).
17. U. Lange, W. Hausner, *Mol. Microbiol.* **52**, 1133 (2004).
18. P. Carter, J. A. Wells, *Science* **237**, 394 (1987).
19. W. Dall'Acqua, P. Carter, *Protein Sci.* **9**, 1 (2000).
20. H. Matsuzaki, G. A. Kassavetis, E. P. Geiduschek, *J. Mol. Biol.* **235**, 1173 (1994).
21. C. Jeon, K. Agarwal, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13677 (1996).
22. J. E. Foster, S. F. Holmes, D. A. Erie, *Cell* **106**, 243 (2001).
23. O. Laptenko, J. Lee, I. Lomakin, S. Borukhov, *EMBO J.* **22**, 6322 (2003).
24. A. M. Poole, D. T. Logan, *Mol. Biol. Evol.* **22**, 1444 (2005).
25. A. Lazcano, J. Fastag, P. Gariglio, C. Ramirez, J. Oro, *J. Mol. Evol.* **27**, 365 (1988).
26. This work is dedicated to the memory of Dmitry Salonin. We thank E. P. Geiduschek for fruitful discussions. This work was supported by NIH grant RO1 GM64530 and a Burroughs Wellcome Career Award (to K.S.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/313/5786/518/DC1
Materials and Methods
SOM Text
Figs. S1 to S3
Table S1
References

14 March 2006; accepted 8 June 2006
10.1126/science.1127422

MC1R Germline Variants Confer Risk for *BRAF*-Mutant Melanoma

Maria Teresa Landi,[1]*† Jürgen Bauer,[2,3]* Ruth M. Pfeiffer,[1] David E. Elder,[4] Benjamin Hulley,[1] Paola Minghetti,[5] Donato Calista,[5] Peter A. Kanetsky,[7] Daniel Pinkel,[6] Boris C. Bastian[2]

Germline variants in *MC1R*, the gene encoding the melanocortin-1 receptor, and sun exposure increase risk for melanoma in Caucasians. The majority of melanomas that occur on skin with little evidence of chronic sun-induced damage (non-CSD melanoma) have mutations in the *BRAF* oncogene, whereas in melanomas on skin with marked CSD (CSD melanoma) these mutations are less frequent. In two independent Caucasian populations, we show that *MC1R* variants are strongly associated with *BRAF* mutations in non-CSD melanomas. In this tumor subtype, the risk for melanoma associated with *MC1R* is due to an increase in risk of developing melanomas with *BRAF* mutations.

Epidemiologic (*1*, *2*) and molecular (*3*, *4*) studies suggest that different types of human melanoma can be distinguished on sun-exposed skin. Tumors on skin with few or no histopathologic signs of CSD, as evidenced by the relative absence of solar elastosis in the surrounding skin, occur in younger individuals and have frequent mutations in the BRAF oncogene (non-CSD melanoma). BRAF encodes a serine/threonine kinase involved in the transduction of mitogenic signals from the cell membrane to the nucleus. By contrast, melanomas on skin with signs of CSD affect older individuals, have different patterns of chromosomal aberrations, and have a lower frequency of BRAF mutations (CSD melanoma) (*4*). Because melanomas on anatomic sites exposed to ultraviolet radiation (UVR) predominantly affect Caucasians, and non-CSD melanomas occur at relatively low UVR doses, we hypothesized that the high frequency of BRAF mutations in this melanoma type is due to a susceptibility factor(s) that occurs at higher frequencies in Caucasian populations (*4*).

A promising candidate susceptibility factor is the melanocortin-1 receptor (MC1R), a G-protein coupled receptor on melanocytes that responds to alpha-melanocyte stimulating hormone (α-MSH) secreted in response to UVR (*5*). The MC1R gene is highly polymorphic in Caucasians (*6*). Its sequence variants can result in partial (r) or complete (R) loss of the receptor's signaling ability, although the degree of functional loss of many MC1R variants is not accurately known. The variants contribute to distinct phenotypic traits such as fair skin, freckling, and red hair (*7*, *8*). Furthermore, MC1R variation has been shown to be a melanoma risk factor (*9*), even beyond its effect on pigmentation (*10*–*12*).

To determine whether there is an association between MC1R variants and BRAF-mutant melanoma, we sequenced the entire coding region of MC1R in germline DNA and the exon 15 of BRAF (where a mutation hot spot is located) in primary cutaneous melanomas from 85 patients from a case-control study conducted in Italy from 1994 to 1999 (*13*, *14*). We performed a similar analysis on an independent set of 112 invasive primary cutaneous melanomas examined at the Department of Dermatology at the University of California, San Francisco, in 2004 and 2005. The MC1R variants identified in the two populations are listed in table S1. The degree of solar elastosis in the skin adjacent to each tumor was assessed independently by two pathologists (*15*) using a multipoint scale from 0 to 3+ (fig. S1). There was good concordance between the two pathologists' scores (weighted kappa = 0.58 and 0.71 for the Italian and U.S. populations, respectively). For statistical analysis, melanomas were classified as non-CSD if they showed only minor signs of solar elastosis (CSD level 0 to 2–) (fig. S1) and as CSD if they had more pronounced solar elastosis (CSD levels 2 to 3+) (fig. S1). As expected, subjects with non-CSD melanomas were younger than those with CSD melanomas, and their tumors arose more frequently on intermittently sun-exposed anatomic sites (e.g., trunk) than on continuously exposed sites (e.g., face) (table S2).

BRAF mutations were more frequent in non-CSD melanoma cases with germline MC1R variants than in those with two wild-type MC1R alleles. When we categorized patients into two groups—homozygous MC1R wild-type versus all others—we found that BRAF mutations were 6 to 13 times as frequent in those with at least one MC1R variant allele compared to those with no MC1R variants (Table 1, upper half). Using a finer MC1R categorization with three groups (zero, one, or two variant alleles), the odds ratio for BRAF mutations in the non-CSD melanomas increased progressively (P = 0.001 and 0.02 for

Table 1. Association between inherited variants of *MC1R* and tumor-specific *BRAF* mutations in non-CSD melanomas. WT, wild type; R, *MC1R* variants with complete loss of function; r, *MC1R* variants with partial loss of function.

	Italy				United States			
MC1R	*BRAF* WT (row %)	*BRAF* mutant (row %)	Odds ratios (95% CI)*	*P*	*BRAF* WT (row %)	*BRAF* mutant (row %)	Odds ratios (95% CI)*	*P*
WT/WT	7 (70.0)	3 (30.0)	Reference		6 (66.7)	3 (33.3)	Reference	
Any variant	9 (19.6)	37 (80.4)	13.2 (2.1–81.4)	0.006	18 (36.7)	31 (63.3)	6.0 (1.2–30.6)	0.03
WT/WT	7 (70.0)	3 (30.0)	Reference		6 (66.7)	3 (33.3)	Reference	
r/WT or R/WT	8 (23.5)	26 (76.5)	10.6 (1.7–67.5)	0.01	15 (44.1)	19 (55.9)	4.1 (0.7–23.0)	0.11
r/r or R/r or R/R	1 (8.3)	11 (91.7)	38.6 (2.5–590.8)	0.009	3 (20.0)	12 (80.0)	10.6 (1.5–74.6)	0.02
Total	16 (28.6)	40 (71.4)		*P* trend = 0.001	24 (41.4)	34 (58.6)		*P* trend = 0.02

*Logistic regression models adjusted by age (quartiles).

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD 20892, USA. [2]Departments of Dermatology and Pathology and Comprehensive Cancer Center, University of California, San Francisco, CA 94143, USA. [3]Department of Dermatology, Eberhard Karls University, 72076 Tübingen, Germany. [4]Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA. [5]Department of Dermatology, M. Bufalini Hospital, Cesena, 47023, Italy. [6]Department of Laboratory Medicine, University of California, San Francisco, CA 94143, USA. [7]Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: landim@mail.nih.gov

Table 2. Melanoma risk in the Italian case-control study by inherited variants of *MC1R* and tumor-specific *BRAF* mutations in non-CSD melanomas. WT, wild type; R, *MC1R* variants with complete loss of function; r, *MC1R* variants with partial loss of function.

MC1R	Controls (No.)	Melanoma cases* (No.)			Odds ratios for melanoma risk (95% CI)†					
		All cases	*BRAF* WT	*BRAF* mutant	All cases	*P*, All cases	*BRAF* WT	*P*, *BRAF* WT	*BRAF* mutant	*P*, *BRAF* mutant
WT/WT	71	10	7	3	Reference		Reference		Reference	
Any variant	100	46	9	37	3.3 (1.5–6.9)	0.002	0.9 (0.3–2.5)	0.79	8.8 (2.6–29.8)	0.0005
WT/WT	71	10	7	3	Reference		Reference		Reference	
r/WT or R/WT	85	34	8	26	2.8 (1.3–6.1)	0.008	1.5 (0.2–13.3)	0.7	7.2 (2.1–24.9)	0.002
r/r or R/r or R/R	15	12	1	11	5.7 (2.1–15.6)	0.001	1.3 (0.2–11.8)	0.8	17.0 (4.2–68.6)	0.0001
Total	171	56	16	40		*P* trend = 0.0003		*P* trend = 0.88		*P* trend < 0.0001

*Only CSD negative cases are included in the analyses. †Logistic regression models adjusted by age (quartiles, in control subjects).

trend in the Italian and U.S. populations, respectively) (Table 1, lower half, and table S3). In an analysis stratified by median age, the association between MC1R and melanoma risk by BRAF mutation status was stronger in the younger subjects (table S4). However, formal tests for interaction between age and MC1R were not significant ($P = 0.22$ and $P = 0.13$ in the Italian and U.S. populations, respectively). MC1R variation had no effect on the frequency of BRAF mutations in melanomas with CSD, although the small number of CSD-positive subjects precluded a formal statistical analysis in the Italian group (table S5).

Comparison of the non-CSD Italian cases with 171 healthy Italian controls showed that the overall melanoma risk was higher by a factor of 3.3 [95% confidence interval (CI) 1.5 to 6.9] in individuals with any MC1R variant allele compared to individuals with no variant alleles and that the risk increased with the number of variant MC1R alleles (Table 2). By stratifying the tumors on the basis of the presence or absence of BRAF mutations, it became evident that the risk was confined to the melanomas with BRAF mutations. The odds ratio increased from 7.2 (95% CI = 2.1 to 24.9) for individuals with one MC1R variant allele to 17.0 (95% CI 4.2 to 68.6) for those with multiple variant alleles when compared with individuals with no MC1R variants ($P < 0.0001$ for trend across categories) (Table 2 and table S6). These results remain significant when using a Bonferroni correction for multiple testing. BRAF mutations were not associated with phenotypic characteristics that are usually associated with sun sensitivity, such as hair color, eye color, spectrophotometrically assessed skin color (*15*), and tanning ability (see table S7 for a comprehensive list).

The relation between BRAF mutations in melanoma and sun exposure is complex and intriguing. On the one hand, sun exposure appears necessary for the development of BRAF mutations because melanomas on mucosa-lined body cavities, the soles, the palms, and subungual sites have low mutation frequencies (11 to 23%) compared to the ~60% mutation frequency in non-CSD melanoma (*4*). On the other hand, melanomas developing in older subjects, after accumulated sun exposure sufficient to produce CSD in the surrounding skin, also exhibit lower BRAF mutation frequencies, arguing against a simple link between UVR exposure and BRAF mutation. Moreover most BRAF mutations do not show the standard C > T signature of direct UVR induction. This paradoxical relationship motivated our hypothesis that there is an inherited susceptibility factor(s) that predisposes individuals to develop BRAF-mutant melanoma under limited sun exposure or earlier in life and that UVR may act indirectly to promote these mutations.

Our results show that variant alleles of MC1R are at least one component of this hypothesized susceptibility. BRAF mutations are a characteristic feature of more than 80% of the non-CSD melanomas in individuals with two variant MC1R alleles but only in ~30% of individuals with wild-type MC1R (Table 1). The mechanism mediating this susceptibility is currently unknown; however, previous studies suggest that it may in part be independent of pigmentation (*10–12*). One possibility is increased generation of reactive oxygen species in carriers of MC1R variants (*16*), which could be independent of pigmentation (*17*) and directly induce the A > T transversion characteristic of the common BRAF V600E mutation in exon 15.

Epidemiological studies often identify associations between cancer risk and environmental exposures, but tumors developing in response to comparable environmental exposure frequently show a variety of somatic changes. Such differences may be due to the stochastic nature of mutation coupled with selection during tumor development. Alternatively, as we show here, the difference may be due to specific inherited genetic variants. Our discovery of the MC1R-BRAF relationship was dependent on careful classification of melanomas into CSD and non-CSD subtypes. We expect that similar subtyping of other cancers will reveal important associations of environmental exposures with germline variants and somatic genetic alterations.

References and Notes

1. D. C. Whiteman *et al.*, *J. Natl. Cancer Inst.* **95**, 806 (2003).
2. V. Siskind, D. C. Whiteman, J. F. Aitken, N. G. Martin, A. C. Green, *Cancer Causes Control* **16**, 193 (2005).
3. J. L. Maldonado *et al.*, *J. Natl. Cancer Inst.* **95**, 1878 (2003).
4. J. A. Curtin *et al.*, *New Eng. J. Med.* **353**, 2135 (2005).
5. V. Chhajlani, J. E. Wikberg, *FEBS Lett.* **309**, 417 (1992).
6. F. Rouzaud, A. L. Kadekaro, Z. A. Abdel-Malek, V. J. Hearing, *Mutat. Res.* **571**, 133 (2005).
7. L. Naysmith *et al.*, *J. Invest. Dermatol.* **122**, 423 (2004).
8. D. L. Duffy *et al.*, *Hum. Mol. Genet.* **13**, 447 (2004).
9. P. Valverde *et al.*, *Hum. Mol. Genet.* **5**, 1663 (1996).
10. J. S. Palmer *et al.*, *Am. J. Hum. Genet.* **66**, 176 (2000).
11. C. Kennedy *et al.*, *J. Invest. Dermatol.* **117**, 294 (2001).
12. M. T. Landi *et al.*, *J. Natl. Cancer Inst.* **97**, 998 (2005).
13. M. T. Landi *et al.*, *Br. J. Cancer* **85**, 1304 (2001).
14. Materials and methods are available as supporting material on *Science* Online.
15. A. V. Brenner, J. H. Lubin, D. Calista, M. T. Landi, *Am. J. Epidemiol.* **156**, 353 (2002).
16. M. C. Scott *et al.*, *J. Cell Sci.* **115**, 2349 (2002).
17. K. Wakamatsu *et al.*, *Pigment Cell Res.* **19**, 154 (2006).
18. We thank A. M. Goldstein and M. Tucker for helpful discussions. This study was supported by the Intramural Research Program of NIH, National Cancer Institute, Division of Cancer Epidemiology and Genetics, and by National Cancer Institute grants RO1 CA5558 to M.T.L., R33 CA95300 and PO1 CA025874-25-A1 to B.C.B., RO1 CA94963 to D.P., K07 CA80700 to P.A.K., and Deutsche Forschungsgemeinschaft stipend BA 2852/1-1 to J.B.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1127515/DC1
Materials and Methods
Tables S1 to S8
Fig. S1

16 March 2006; accepted 8 June 2006
Published online 29 June 2006;
10.1126/science.1127515
Include this information when citing this paper.

Chimpanzee Reservoirs of Pandemic and Nonpandemic HIV-1

Brandon F. Keele,[1] Fran Van Heuverswyn,[2] Yingying Li,[1] Elizabeth Bailes,[3] Jun Takehisa,[1] Mario L. Santiago,[1]* Frederic Bibollet-Ruche,[1] Yalu Chen,[1] Louise V. Wain,[3] Florian Liegeois,[2] Severin Loul,[4] Eitel Mpoudi Ngole,[4] Yanga Bienvenue,[4] Eric Delaporte,[2] John F. Y. Brookfield,[3] Paul M. Sharp,[3] George M. Shaw,[1,5] Martine Peeters,[2] Beatrice H. Hahn[1]†

Human immunodeficiency virus type 1 (HIV-1), the cause of human acquired immunodeficiency syndrome (AIDS), is a zoonotic infection of staggering proportions and social impact. Yet uncertainty persists regarding its natural reservoir. The virus most closely related to HIV-1 is a simian immunodeficiency virus (SIV) thus far identified only in captive members of the chimpanzee subspecies *Pan troglodytes troglodytes*. Here we report the detection of SIVcpz antibodies and nucleic acids in fecal samples from wild-living *P. t. troglodytes* apes in southern Cameroon, where prevalence rates in some communities reached 29 to 35%. By sequence analysis of endemic SIVcpz strains, we could trace the origins of pandemic (group M) and nonpandemic (group N) HIV-1 to distinct, geographically isolated chimpanzee communities. These findings establish *P. t. troglodytes* as a natural reservoir of HIV-1.

Since the first detection of an HIV-1–related lentivirus in chimpanzees (*1*, *2*), this species has been suspected as the source of the human AIDS pandemic. However, a crucial missing link in the chain of evidence implicating SIVcpz in the origin of HIV-1 and AIDS has been the absence of a recognizable virus reservoir in wild-living apes. Chimpanzees (*Pan troglodytes*) are classified into four subspecies on the basis of differences in mitochondrial DNA sequence (*3*): *P. t. verus* in west Africa; *P. t. vellerosus* in Nigeria and northern Cameroon; *P. t. troglodytes* in southern Cameroon, Gabon, and the Republic of Congo; and *P. t. schweinfurthii* in the Democratic Republic of Congo and countries to the east (Fig. 1). Two of these subspecies, *P. t. troglodytes* and *P. t. schweinfurthii*, are known to harbor SIVcpz, and their viruses form divergent subspecies-specific phylogenetic lineages (SIVcpz*Ptt* and SIVcpz*Pts*) (*4*). HIV-1 is most closely related to SIVcpz*Ptt* (*5*), but this virus has been detected only rarely and then only in captive apes (*1*, *5*–*7*). There is no counterpart of SIVcpz*Pts* that is known to infect humans (*4*, *8*–*10*).

Wild-living chimpanzees are reclusive and highly endangered and live in remote jungle areas. To study chimpanzees in their natural habitat, we developed methods to detect SIVcpz-specific antibodies and nucleic acids in fecal samples collected from the forest floor (*9*–*11*). In addition, we developed genotyping approaches to amplify host mitochondrial and genomic markers (polymorphic microsatellite loci) from these same specimens for species, gender, and individual identification (*11*, *12*). These methods were validated in captive and habituated apes of known infection status (*13*). We used these noninvasive approaches to conduct the first molecular epidemiological field study of SIVcpz in wild-living nonhabituated chimpanzees in west central Africa.

Cameroon is home to two chimpanzee subspecies, *P. t. vellerosus* in the north and *P. t. troglodytes* in the south, with the Sanaga River forming the boundary between their ranges (Fig. 1). In the present study, we collected 599 fecal specimens at 10 forest sites throughout the southern part of Cameroon (Fig. 1). All field sites, except one (WE), were in the range of the *P. t. troglodytes* subspecies. To establish the species and subspecies origin of each sample, a 498–base pair (bp) mitochondrial DNA (mtDNA) (D-loop) fragment was amplified from fecal DNA and subjected to phylogenetic analysis (*13*). Eighty-six specimens were degraded, and 67 samples contained gorilla mtDNA sequences (table S1). The remaining 446 samples were of chimpanzee origin: 423 from *P. t. troglodytes* and 23 from *P. t. vellerosus*. These comprised 82 unique mtDNA haplotypes (fig. S1 and table S2). Consistent with the recognized ranges of the two subspecies, all 23 *P. t. vellerosus* speci-

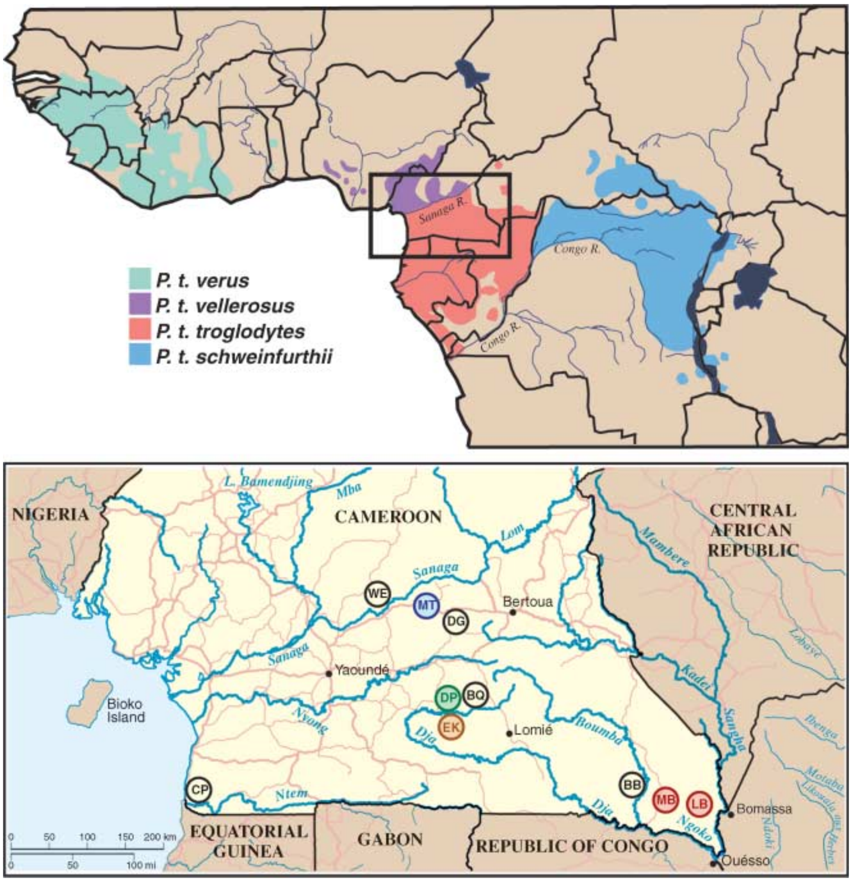


Fig. 1. Natural ranges of the four chimpanzee subspecies (top) and locations of wild chimpanzee study sites WE, MT, DG, DP, BQ, EK, CP, BB, MB, and LB in southern Cameroon (inset and bottom). Field sites with endemic SIVcpz*Ptt* infection are color-coded to correspond with the SIVcpz*Ptt* lineages shown in Figs. 3 and 4.

[1] Departments of Medicine and Microbiology, University of Alabama at Birmingham, Birmingham, AL, USA. [2] Laboratoire Retrovirus, UMR145, Institut de Recherche pour le Développement and Department of International Health, University of Montpellier I, 911 Avenue Agropolis, Boite Postale 64501, 34394 Montpellier Cedex 5, France. [3] Institute of Genetics, University of Nottingham, Queens Medical Centre, Nottingham, NG7 2UH, UK. [4] Projet Prevention du Sida au Cameroun (PRESICA), Boite Postale 1857, Yaoundé, Cameroun. [5] Howard Hughes Medical Institute, 720 South 20th Street, KAUL 816, Birmingham, AL 35294, USA.

*Present address: Gladstone Institute for Virology and Immunology, University of California at San Francisco, 1650 Owens Street, San Francisco, CA 94158, USA.

†To whom correspondence should be addressed. E-mail: bhahn@uab.edu

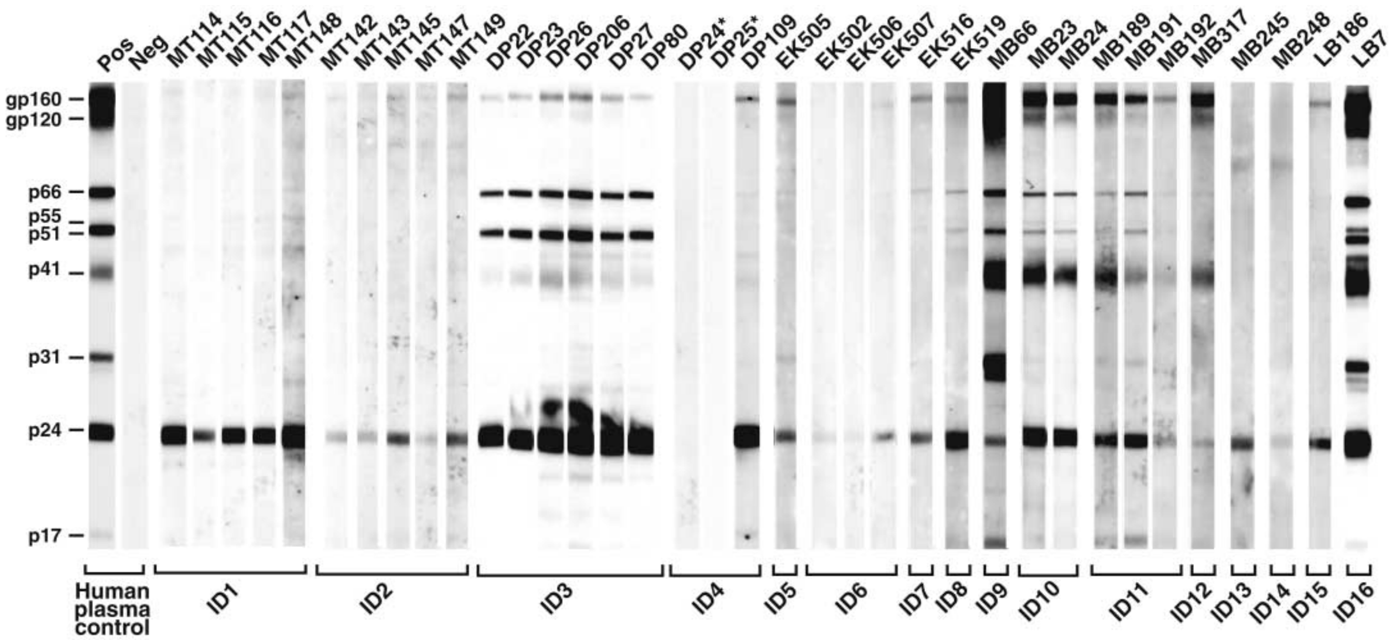


Fig. 2. Detection of SIVcpz antibodies in chimpanzee fecal samples. Fecal samples from wild-living chimpanzees were tested by enhanced chemiluminescent Western blot using HIV-1 antigen–containing strips. Samples are numbered, with letters indicating their collection site as shown in Fig. 1. Samples from the same individual (ID) are grouped. Asterisks indicate two antibody-negative but virion RNA–positive samples (also see table S3). Molecular weights of HIV-1 proteins are indicated. The banding patterns of plasma from HIV-1–infected (Pos) and –uninfected (Neg) humans are shown as controls.

mens were collected north of the Sanaga River, whereas 421 of 423 *P. t. troglodytes* samples were collected south of the river (table S1).

All mtDNA-positive fecal samples were tested for virus-specific antibodies with a sensitive immunoblot assay specifically developed for surveys at remote field sites (*13*). This analysis identified 34 specimens, all from *P. t. troglodytes* apes, that contained antibodies reactive with HIV-1 antigens (Fig. 2). Twelve samples exhibited a strong and broadly cross-reactive Western blot profile that was virtually indistinguishable from the HIV-1–positive human plasma control. Eighteen additional samples reacted with both the HIV-1 envelope (gp160) and major core (p24) proteins, thus also meeting formal criteria for HIV-1/SIVcpz antibody positivity. Four samples (EK502, EK506, MB245, and MB248) reacted only faintly with a single HIV-1 protein (p24) and were classified as indeterminant. None of 23 *P. t. vellerosus* or 67 gorilla specimens exhibited detectable Western blot reactivity to any HIV-1 protein (table S1).

To corroborate the fecal antibody results, RNA was extracted from all immunoblot-reactive samples and subjected to reverse transcription polymerase chain reaction amplification using consensus *env* and *pol* primers. In addition, fecal DNA was used to amplify polymorphic microsatellite loci to identify and distinguish individual apes and to amplify a portion of the amelogenin gene for gender determination (*13*). These analyses revealed that the 34 immunoblot-reactive samples represented 16 different *P. t. troglodytes* apes (7 males and 9 females). Each of these apes had detectable virion RNA in one or more fecal samples (table S3). SIVcpz *env* (~390 bp) and/or *pol* (~890 bp) sequences were amplified from 31 of 34 (91%) immunoblot-reactive samples, including all four specimens with indeterminant Western blot reactivity (Fig. 3 and table S3). These data,

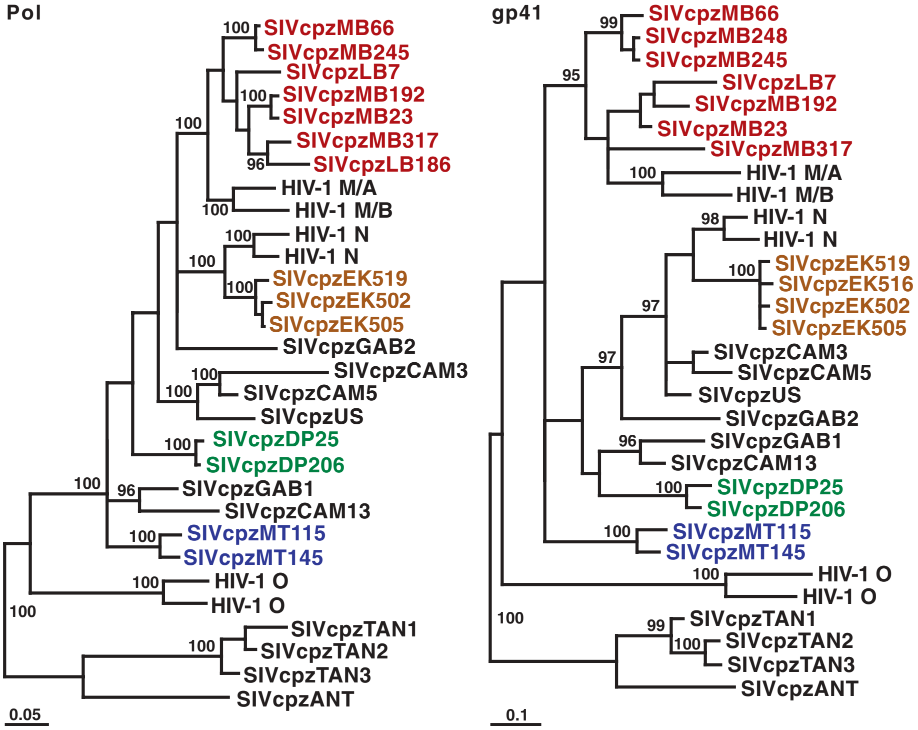


Fig. 3. Phylogenetic analysis of SIVcpz*Ptt* strains from wild *P. t. troglodytes* apes. Newly identified SIVcpz*Ptt* strains are highlighted by colors reflecting their collection sites (Fig. 1). Representative strains of HIV-1 groups M, N, and O and SIVcpz*Pts* (TAN1, TAN2, TAN3, and ANT) are shown. Trees were inferred by the Bayesian method; numbers on nodes are percentage posterior probabilities (only values above 95% are shown). The scale bars represent 0.05 and 0.1 substitutions per site. Pol, polymerase; gp41, envelope transmembrane protein.

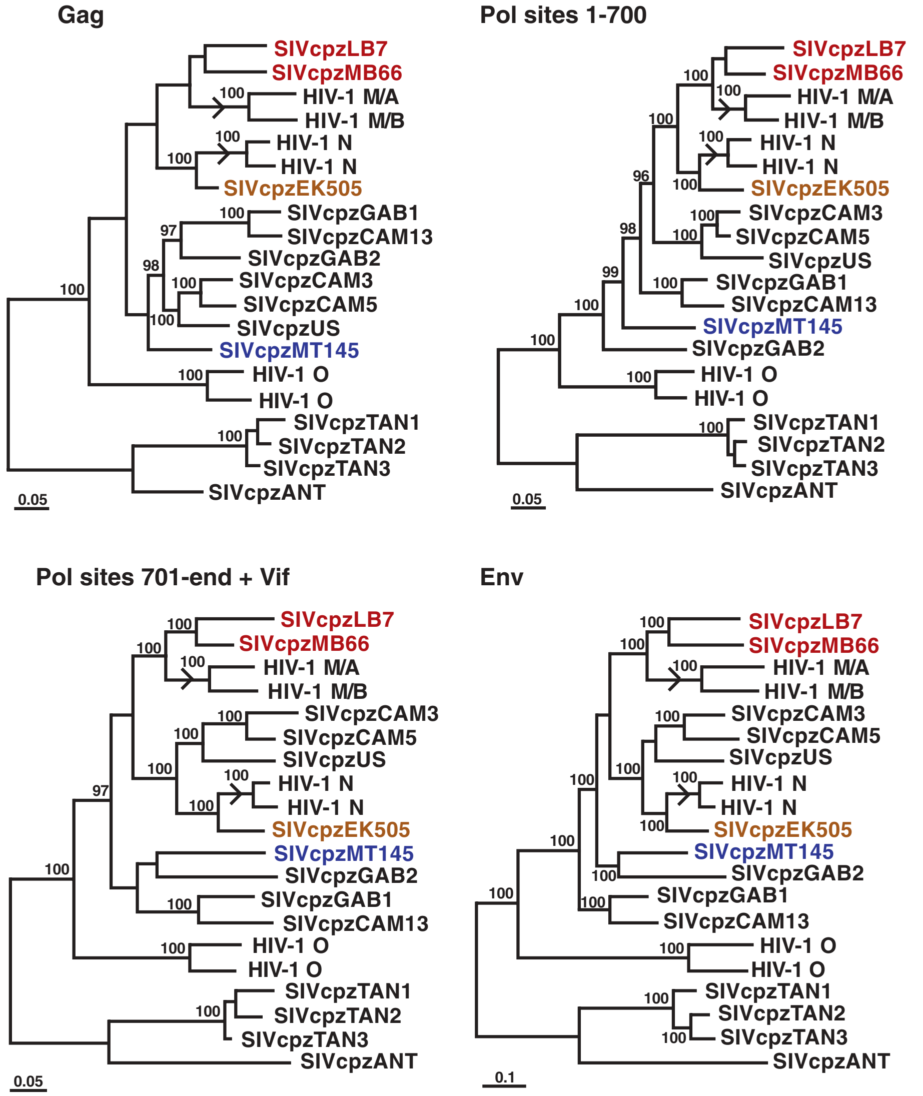


Fig. 4. Evolutionary relationships of SIVcpz and HIV-1 strains based on full-length sequences. Trees were inferred by the Bayesian method for Gag, Pol, and Env; the Pol protein was separated into two fragments at a recombination breakpoint previously identified in HIV-1 group N (*5*, *21*); the C terminus of Pol was concatenated with downstream Vif sequences. Sequences are color-coded as in Fig. 3. Numbers on internal branches indicate estimated posterior probabilities (only values above 95% are shown). The scale bars represent 0.05 and 0.1 substitutions per site. Arrows indicate branches where cross-species transmissions gave rise to HIV-1 groups M and N.

together with previous findings for SIVcpz*Pts*-infected apes (*10*), indicate that fecal antibody reactivity to a single HIV-1 Gag protein is indicative of SIVcpz*Ptt* infection (*14*).

The prevalence of SIVcpz*Ptt* infection in wild chimpanzee communities was estimated for each of the 10 field sites (table S1). For the DP, EK, MB, BB, and LB communities, this was done based on the proportion of infected individuals as determined by microsatellite analyses, taking into consideration assay sensitivities and specimen degradation (tables S1 and S4). For the remaining sites, prevalence rates were estimated based on the proportion of antibody- and/or SIVcpz virion RNA–positive fecal samples, while also adjusting for repeat sampling (*13*). The results indicated widespread but notably uneven SIVcpz*Ptt* infection of wild-living *P. t. troglodytes* apes, with prevalence rates ranging from 23 to 35% in the LB, EK, and MB communities; 4 to 5% in the DP and MT communities; and the absence of infection in the WE, DG, BQ, BB, and CP communities.

To determine the evolutionary relationships of the 16 new SIVcpz*Ptt* viruses to each other and to previously characterized SIVcpz and HIV-1 strains, *pol* and *env* sequences were subjected to phylogenetic analyses. All of the newly identified SIVcpz strains were found to fall within the radiation of SIVcpz*Ptt* strains from captive *P. t. troglodytes* apes, which also includes HIV-1 groups M (pandemic) and N (nonpandemic) but not group O or SIVcpz*Pts* (Fig. 3). The new *P. t. troglodytes* viruses exhibited significant phylogeographic clustering: SIVcpz sequences from the EK, DP, MT, and MB/LB collection sites formed well-separated clades corresponding to their field site of origin. One of these clades included closely related SIVcpz strains (EK519, EK516, EK502, and EK505), probably reflecting recent virus transmission within that community. The remaining clades were each composed of more divergent but still monophyletic SIVcpz strains (Fig. 3). Thus, chimpanzee populations separated by long distances or major geographical barriers such as rivers (Fig. 1) harbored distinct SIVcpz lineages (such as populations EK, DP, and MT), whereas neighboring communities not separated by such barriers harbored viruses that were phylogenetically interspersed (such as populations MB and LB).

The phylogeographic clustering of the newly identified SIVcpz*Ptt* strains allowed us to trace the origins of present-day human AIDS viruses to distinct chimpanzee communities. In subgenomic *pol* and *env* regions, SIVcpz*Ptt* strains from the MB/LB and EK sites were much more closely related to HIV-1 groups M and N, respectively, than were any previously identified SIVcpz strains (Fig. 3). Full-length genome analysis of 4 of the 16 new viruses confirmed and extended these findings, revealing strong statistical support for the clustering of HIV-1 groups M and N with the MB/LB and EK lineages of SIVcpz*Ptt*, respectively (Fig. 4). Moreover, inclusion of the new viruses reduced the lengths of the branches marking the cross-species transmission events for all genomic regions by almost half (arrows in Fig. 4). Given these short branch lengths, it is highly unlikely that other SIVcpz*Ptt* strains exist that are significantly more closely related to HIV-1 groups M and N than are the viruses from the MB/LB and EK communities. Indeed, expanded field studies in southern Cameroon by our group have identified additional SIVcpz*Ptt* strains, including nine from the MB/LB area, whose sequences support this conclusion and corroborate the phylogenetic relationships shown in Figs. 3 and 4 (*15*). Thus, an extensive set of molecular epidemiological data all points to chimpanzees in southeastern and south central Cameroon as the sources of HIV-1 groups M and N, respectively.

The findings presented here, together with prior studies, provide for the first time a clear picture of the origin of HIV-1 and the seeds of the AIDS pandemic. SIVcpz, the progenitor of HIV-1, arose as a recombinant of ancestors of SIV lineages presently infecting red-capped mangabeys and *Cercopithecus* monkeys in west-central Africa (*16*). Chimpanzees acquired this recombinant virus, or its progenitors, by cross-species transmission some time after the split of *P. t. verus* and *P. t. vellerosus* from the other subspecies (fig. S1) but possibly before the divergence of *P. t. schweinfurthii* from *P. t. troglodytes* (*4*). This explains the absence of SIVcpz infection in present-day

P. t. verus and *P. t. vellerosus* apes, the presence of SIVcpz infection in *P. t. troglodytes* and *P. t. schweinfurthii* apes, and the phylogenetic separation of SIVcpz*Ptt* from SIVcpz*Pts* viruses (*4*, *7*, *9*, *15*). HIV-1 groups M, N, and O each resulted from independent cross-species transmissions of SIVcpz*Ptt* from *P. t. troglodytes* to humans early in the 20th century (*17–19*). We show here that the SIVcpz*Ptt* strain that gave rise to HIV-1 group M belonged to a viral lineage that persists today in *P. t. troglodytes* apes in southeastern Cameroon. That virus was probably transmitted locally. From there it appears to have made its way via the Sangha River (or other tributaries) south to the Congo River and on to Kinshasa where the group M pandemic was probably spawned (*20*). HIV-1 group N, which has been identified in only a small number of AIDS patients from Cameroon (*21*, *22*), derived from a second SIVcpz*Ptt* lineage in south central Cameroon and remained geographically more restricted. The source of HIV-1 group O remains unknown but will probably yield to further study of wild ape populations not yet sampled. Given the extensive genetic diversity and phylogeographic clustering of SIVcpz now recognized, and the vast areas of west central Africa not yet sampled (Fig. 1), it is quite possible that still other SIVcpz lineages exist that could pose risks of human infection and prove problematic for HIV diagnostics and vaccines. The present report describes molecular tools and noninvasive strategies that can be used to explore these possibilities as well as the molecular ecology of pathogens in endangered species more generally.

References and Notes

1. M. Peeters *et al.*, *AIDS* **3**, 625 (1989).
2. T. Huet *et al.*, *Nature* **345**, 356 (1990).
3. C. P. Groves, in *Mammalian Species of the World: A Taxonomic and Geographic Reference*, D. E. Wilson, D. M. Reader, Eds. (Smithsonian Institution Press, Washington, DC, ed. 2, 1993), pp. 243–277.
4. P. M. Sharp, G. M. Shaw, B. H. Hahn, *J. Virol.* **79**, 3891 (2005).
5. F. Gao *et al.*, *Nature* **397**, 436 (1999).
6. S. Corbet *et al.*, *J. Virol.* **74**, 529 (2000).
7. E. Nerrienet *et al.*, *J. Virol.* **79**, 1312 (2005).
8. M. M. Vanden Haesevelde *et al.*, *Virology* **221**, 346 (1996).
9. M. L. Santiago *et al.*, *Science* **295**, 465 (2002).
10. M. Worobey *et al.*, *Nature* **428**, 820 (2004).
11. M. L. Santiago *et al.*, *J. Virol.* **77**, 7545 (2003).
12. M. L. Santiago *et al.*, *J. Virol.* **79**, 12515 (2005).
13. Materials and methods are available as supporting material on *Science* Online.
14. In contrast to plasma samples from uninfected humans, which exhibit false positive Western blot reactivity to HIV-1 p24 in as many as 10 to 15% of individuals (www.fda.gov/cber/products/hiv1cam052898.htm), we have found no such nonspecific cross-reactivity of chimpanzee immunoglobulin extracted by the RNA*later* method from over 2000 fecal specimens.
15. F. Van Heuverswyn *et al.*, *13th Conference on Retroviruses and Opportunistic Infections*, Abstract **132**, available at www.retroconference.org/2006/Abstracts/26513.htm (2006).
16. E. Bailes *et al.*, *Science* **300**, 1713 (2003).
17. B. H. Hahn, G. M. Shaw, K. M. De Cock, P. M. Sharp, *Science* **287**, 607 (2000).
18. B. T. K. Korber *et al.*, *Science* **288**, 1789 (2000).
19. P. M. Sharp *et al.*, *Philos. Trans. R. Soc. London Ser. B* **356**, 867 (2001).
20. N. Vidal *et al.*, *J. Virol.* **74**, 10498 (2000).
21. F. Simon *et al.*, *Nat. Med.* **4**, 1032 (1998).
22. J. Yamaguchi *et al.*, *AIDS Res. Hum. Retroviruses* **22**, 83 (2006).
23. We thank the Cameroonian Ministries of Health, Environment and Forestry, and Research for permission to perform this study; J. Dupain, C. Neel, and M. Epanda (Projet Grands Singes); L. Usongo, D. Dontego, F. Espedi, and B. Tshikangwa (World Wildlife Fund); and G. Etoga and D. M'bohand (Ministry of Environment and Forestry) for assistance in the field. This work was supported in part by the NIH (grants R01 AI50529, R01 AI58715, and P30 AI 27767), the Bristol Myers Freedom to Discover Program, the Institut de Recherche pour le Développement, and the Howard Hughes Medical Institute. New SIVcpz*Ptt* sequences are available at GenBank under accession numbers DQ370366-DQ370419 and DQ373063-DQ373066; chimpanzee mtDNA sequences are available under DQ367532-DQ367613 and DQ370307-DQ370365.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1126531/DC1
Materials and Methods
Fig. S1
Tables S1 to S4
References

22 February 2006; accepted 17 May 2006
Published online 25 May 2006;
10.1126/science.1126531
Include this information when citing this paper.

Mast Cells Can Enhance Resistance to Snake and Honeybee Venoms

Martin Metz,[1] Adrian M. Piliponsky,[1] Ching-Cheng Chen,[1] Verena Lammel,[1] Magnus Åbrink,[2] Gunnar Pejler,[2] Mindy Tsai,[1] Stephen J. Galli[1]*

Snake or honeybee envenomation can cause substantial morbidity and mortality, and it has been proposed that the activation of mast cells by snake or insect venoms can contribute to these effects. We show, in contrast, that mast cells can significantly reduce snake-venom–induced pathology in mice, at least in part by releasing carboxypeptidase A and possibly other proteases, which can degrade venom components. Mast cells also significantly reduced the morbidity and mortality induced by honeybee venom. These findings identify a new biological function for mast cells in enhancing resistance to the morbidity and mortality induced by animal venoms.

Venomous reptiles and their prey have coexisted for $\sim$200 million years (*1*), and snake envenomation still accounts for considerable human morbidity and mortality worldwide (*2*, *3*) (SOM Text 1). The mechanisms by which snake envenomation can produce tissue injury and death have been studied extensively (*3–5*), and it is known that many components of snake venoms can induce mammalian mast cells (MCs) to release potent biologically active mediators (*6*, *7*). These MC products in turn can promote an increase in vascular permeability, local inflammation, abnormalities of the clotting and fibrinolysis systems, and shock (*8*, *9*).

Accordingly, it has been considered that the activation of tissue MCs can contribute importantly to the local tissue injury, systemic distribution of venom components, and death associated with snake envenomation (*6*, *7*). This hypothesis is consistent with the well-understood role of MCs in the pathology of allergic disorders such as anaphylaxis and asthma (*8–11*). However, MCs can enhance survival in certain models of innate immunity to bacterial infection (*12–15*). In one such model, MCs can reduce morbidity and mortality in part by promoting the degradation of the potent endogenous vasoconstrictor peptide endothelin-1 (ET-1) (*16*). The most toxic components of the venom of *Atractaspis engaddensis* (the burrowing asp or Israeli mole viper) are the sarafotoxins, which exhibit a very high homology ($\sim$70% at the amino acid level) to ET-1 (*17*).

When various amounts of *A. engaddensis* venom (*A.e.*v.) were administered intraperitoneally, wild-type mice developed significant reductions in body temperature at a dose of 5 μg, and death occurred at 50 μg (fig. S1). By contrast, as little as 5 μg of *A.e.*v. induced death in $Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$ mice, which are genetically deficient in MCs (*18*). Levels of sarafotoxins in the peritoneal cavity of wild-type mice were significantly lower than those in the corresponding $Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$ mice at all amounts of *A.e.*v. tested that were ≥5 μg (fig. S1). Although intraperitoneal injection has been recommended for analyses of the systemic toxicity of snake venoms (*4*), many snake bites are to the skin and subcutaneous tissue. MC-deficient mice were also much more susceptible than wild-type mice to the development of hypothermia and death when *A.e.*v. (10 μg) was injected subcutaneously (fig. S2).

*A.e.*v. contains several toxic compounds, including sarafotoxins 6a, 6b, 6c, and 6d, and hemorrhagins, but the most toxic of these is

[1] Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305–5324, USA. [2] Department of Molecular Biosciences, Swedish University of Agricultural Sciences, The Biomedical Center, Box 575, 751 23 Uppsala, Sweden.

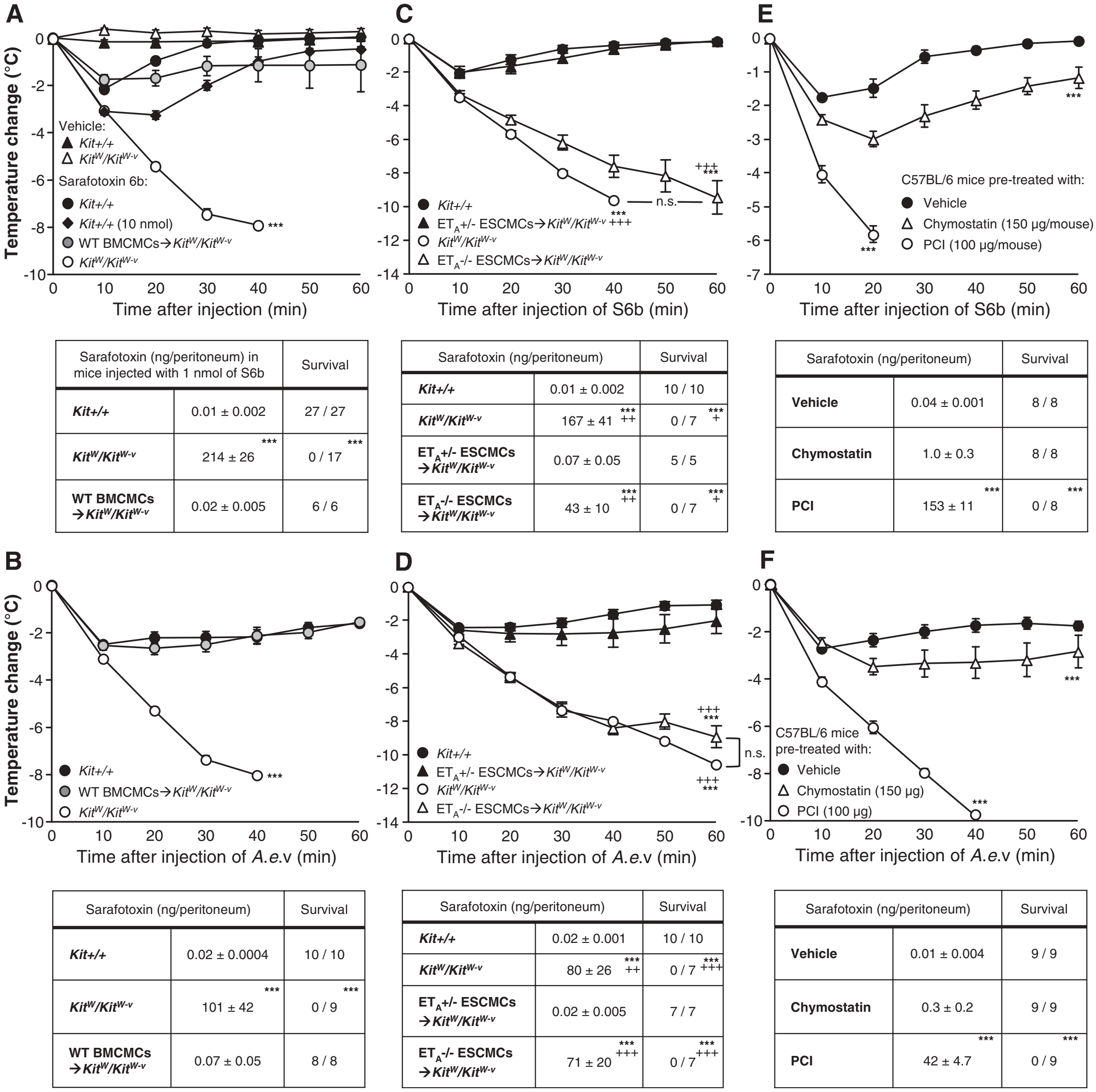
*To whom correspondence should be addressed. E-mail: sgalli@stanford.edu

sarafotoxin 6b (S6b) (*19*). When injected intraperitoneally with 1 nmol of S6b, each MC-deficient $Kit^{W}/Kit^{W\text{-}v}$ mouse developed severe hypothermia and died within 1 hour, whereas the wild-type $Kit^{+/+}$ mice developed only a slight drop in temperature and recovered completely, even when injected with 10 times the amount of S6b (Fig. 1A). Injection of unfractionated venom produced similar results (Fig. 1B). Similar results also were obtained when C57BL/6 and MC-deficient $Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$ were injected intraperitoneally with S6b or *A.e.*v. (fig. S3). Intraperitoneal injection of S6b or *A.e.*v. also induced extensive degranulation of peritoneal MCs (PMCs) (fig. S4, A and B), a finding readily detected at 10 min after injection of *A.e.*v. (fig. S4), and levels of sarafotoxins were almost undetectable in the peritoneal cavity of $Kit^{+/+}$ mice 60 min after S6b or *A.e.*v. injection, whereas the peritoneal cavity of $Kit^{W}/Kit^{W\text{-}v}$ mice contained high levels of sarafotoxins (Fig. 1, A and B).

To assess whether the differences in the responses of $Kit^{+/+}$ versus $Kit^{W}/Kit^{W\text{-}v}$ mice specifically reflected the lack of MCs in the $Kit^{W}/Kit^{W\text{-}v}$ mice, we examined $Kit^{W}/Kit^{W\text{-}v}$ mice that had been engrafted with bone marrow–derived cultured MCs (BMCMCs) derived from the $Kit^{+/+}$ mice (wild-type BMCMCs→$Kit^{W}/Kit^{W\text{-}v}$ mice) (*15*, *20*). Upon injection of S6b or *A.e.*v., MC-engrafted $Kit^{W}/Kit^{W\text{-}v}$ mice exhibited very low sarafotoxin levels in the peritoneal cavity and were protected against hypothermia and death (Fig. 1, A and B); similar results were obtained with wild-type MC-engrafted $Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$ mice (fig. S3).

Thus, MCs can significantly reduce levels of sarafotoxins in the peritoneal cavity after intraperitoneal injection of S6b or *A.e.*v. in vivo and can markedly limit the systemic toxicity and death induced by either S6b or unfractionated *A.e.*v.

S6b, like ET-1, activates the ET_A receptor, which is expressed by MCs (*16*). We generated two groups of mice that contained MCs that either expressed or lacked ET_A. Because $ET_A{}^{-/-}$ mice are not viable, we generated MCs from $ET_A{}^{+/-}$ or $ET_A{}^{-/-}$ embryonic stem (ES) cells (*21*) and then adoptively transferred these ES cell–derived cultured MCs (ESCMCs) into $Kit^{W}/Kit^{W\text{-}v}$ mice (Fig. 1, C and D). Although $ET_A{}^{+/-}$ and $ET_A{}^{-/-}$ ESCMC-engrafted mice contain similar numbers of MCs in the peritoneum (*16*), those that received $ET_A{}^{-/-}$ ESCMCs developed severe hypo-



Sarafotoxin (ng/peritoneum) in mice injected with 1 nmol of S6b		Survival
Kit+/+	0.01 ± 0.002	27 / 27
Kit^W/Kit^W-v	214 ± 26 ***	0 / 17 ***
WT BMCMCs →*Kit^W/Kit^W-v*	0.02 ± 0.005	6 / 6

Sarafotoxin (ng/peritoneum)		Survival
Kit+/+	0.01 ± 0.002	10 / 10
Kit^W/Kit^W-v	167 ± 41 *** ++	0 / 7 *** +
ET_A+/- ESCMCs →*Kit^W/Kit^W-v*	0.07 ± 0.05	5 / 5
ET_A-/- ESCMCs →*Kit^W/Kit^W-v*	43 ± 10 *** ++	0 / 7 *** +

Sarafotoxin (ng/peritoneum)		Survival
Vehicle	0.04 ± 0.001	8 / 8
Chymostatin	1.0 ± 0.3	8 / 8
PCI	153 ± 11 ***	0 / 8 ***

Sarafotoxin (ng/peritoneum)		Survival
Kit+/+	0.02 ± 0.0004	10 / 10
Kit^W/Kit^W-v	101 ± 42 ***	0 / 9 ***
WT BMCMCs →*Kit^W/Kit^W-v*	0.07 ± 0.05	8 / 8

Sarafotoxin (ng/peritoneum)		Survival
Kit+/+	0.02 ± 0.001	10 / 10
Kit^W/Kit^W-v	80 ± 26 *** ++	0 / 7 *** +++
ET_A+/- ESCMCs →*Kit^W/Kit^W-v*	0.02 ± 0.005	7 / 7
ET_A-/- ESCMCs →*Kit^W/Kit^W-v*	71 ± 20 *** +++	0 / 7 *** +++

Sarafotoxin (ng/peritoneum)		Survival
Vehicle	0.01 ± 0.004	9 / 9
Chymostatin	0.3 ± 0.2	9 / 9
PCI	42 ± 4.7 ***	0 / 9 ***

Fig. 1. MCs can protect mice from *A.e.*v.- and S6b-induced morbidity and mortality. Shown are changes in rectal temperatures, levels of sarafotoxins in the peritoneal lavage fluid (at time of death or at 60 min, whichever was sooner), and 24-hour survival after intraperitoneal injection of S6b, *A.e.*v., or saline. (**A**, **C**, and **E**) Results after injection of 1 nmol or [in (A) only] 10 nmol of S6b in 250 μl saline (0.9% sterile NaCl) or 250 μl saline alone. In (A), survival and sarafotoxin data in the table are only for mice injected with 1 nmol S6b; survival in all other groups was 100%. (**B**, **D**, and **F**) Results after injection of *A.e.*v. [10 μg in 250 μl phosphate-buffered saline (PBS)]. Subjects were (A and B) $Kit^{+/+}$ mice, MC-deficient $Kit^{W}/Kit^{W\text{-}v}$ mice, and $Kit^{W}/Kit^{W\text{-}v}$ mice that had been engrafted intraperitoneally with MCs derived in vitro from the bone marrow of $Kit^{+/+}$ mice (WT BMCMCs→$Kit^{W}/Kit^{W\text{-}v}$); (C and D) $Kit^{+/+}$ mice, MC-deficient $Kit^{W}/Kit^{W\text{-}v}$ mice, and $Kit^{W}/Kit^{W\text{-}v}$ mice that had been engrafted intraperitoneally with ES cell–derived cultured mast cells (ESCMCs) that did ($ET_A{}^{+/-}$) or did not ($ET_A{}^{-/-}$) express ET_A; and (E and F) C57BL/6 mice that had received intraperitoneal injections of vehicle (saline, 300 μl), chymostatin (150 μg in 300 μl of saline), or PCI (100 μg in 300 μl of saline) 10 min before injection of S6b or saline. (A and B) ***$P < 0.005$ versus corresponding values for $Kit^{+/+}$ or WT BMCMCs→$Kit^{W}/Kit^{W\text{-}v}$ mice. (C and D) ***$P < 0.005$ versus $Kit^{+/+}$ mice; $+P < 0.05$, $++P < 0.01$, or $+++P < 0.005$ versus $ET_A{}^{+/-}$ ESCMCs→$Kit^{W}/Kit^{W\text{-}v}$ mice; n.s., not significant ($P > 0.05$). (E and F) ***$P < 0.005$ versus vehicle.

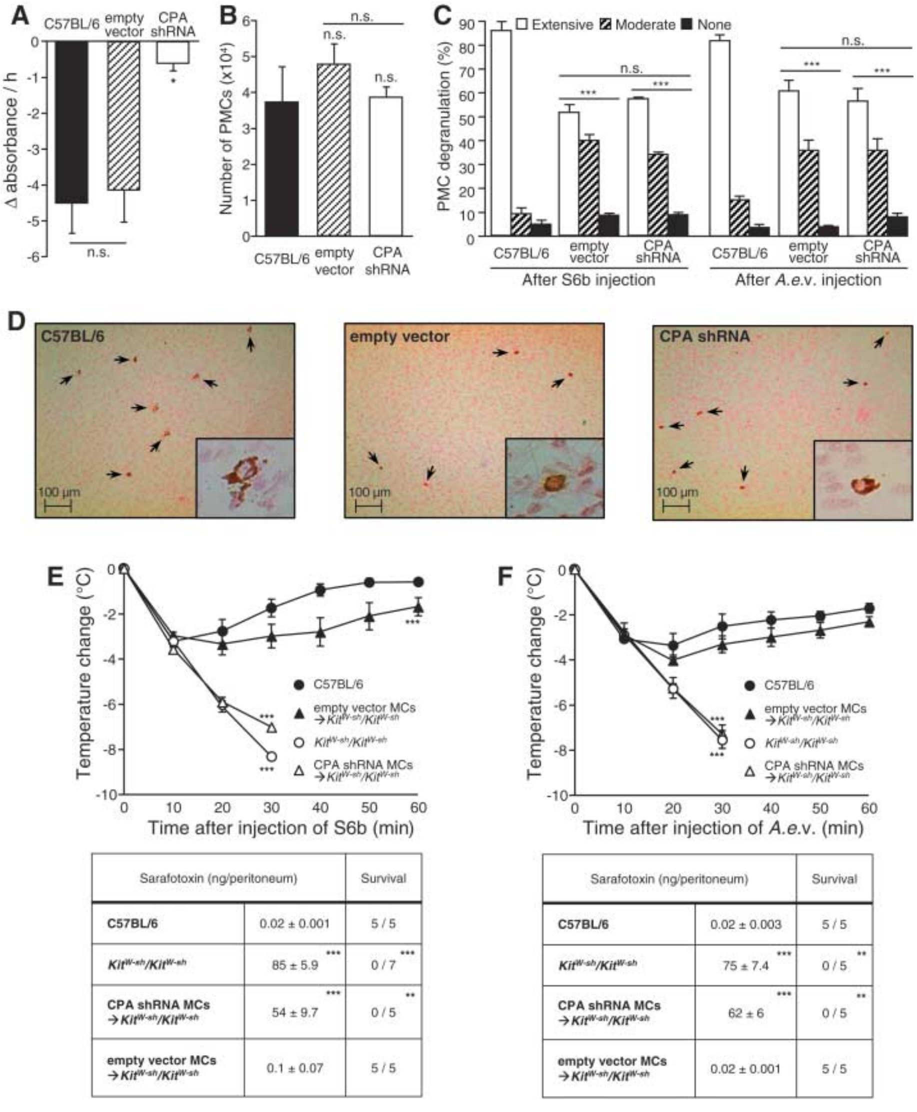
thermia (like that in $Kit^{W}/Kit^{W\text{-}v}$ mice) and died rapidly after intraperitoneal injection of S6b or *A.e.*v. (Fig. 1, C and D). Consistent with their survival responses, $ET_A^{+/-}$ ESCMCs→$Kit^{W}/Kit^{W\text{-}v}$ mice, like $Kit^{+/+}$ mice, exhibited extensive peritoneal mast cell (PMC) degranulation (fig. S4, C and D) and strongly reduced sarafotoxin levels in the peritoneal cavity (Fig. 1, C and D). In contrast, $ET_A^{-/-}$ ESCMCs→$Kit^{W}/Kit^{W\text{-}v}$ mice exhibited significantly lower levels of MC degranulation (fig. S4, C and D) and much higher intraperitoneal levels of sarafotoxins (Fig. 1, C and D).

To assess potential mechanisms by which MCs might reduce venom toxicity, we tested inhibitors of two candidate MC-derived mediators, chymase, an endopeptidase (*16*), and carboxypeptidase A (CPA), an exopeptidase (*22*) [i.e., chymostatin and potato carboxypeptidase inhibitor (PCI), respectively)]. Wild-type C57BL/6 mice pretreated intraperitoneally with PCI developed severe hypothermia and died within 1 hour of S6b or *A.e.*v. injection (Fig. 1, E and F). The extent of PMC degranulation did not differ in wild-type mice pretreated with PCI, vehicle, or chymostatin (fig. S4, E and F). However, sarafotoxin levels in the peritoneum of PCI-pretreated wild-type mice were much higher than those in the vehicle- or chymostatin-pretreated wild-type mice (Fig. 1, E and F), although lower than those in S6b- or *A.e.*v.-injected MC-deficient mice (Fig. 1, A and B). By contrast, compared with the effect of vehicle alone, intraperitoneal pretreatment with chymostatin resulted in only a small reduction in body temperature and a small increase in intraperitoneal levels of sarafotoxins (Fig. 1, E and F). PCI also markedly inhibited the ability of mouse peritoneal cells to degrade S6b ex vivo, whereas chymostatin had little or no effect (fig. S5).

These results were unexpected, because our previous pharmacological experiments have suggested an important role for MC-associated chymase in the degradation of ET-1 (*16*). To examine this further, we undertook a genetic approach to distinguish the relative roles of chymase versus CPA in limiting *A.e.*v toxicity. The only protease with chymotryptic activity in mouse PMCs is mouse mast cell protease-4 (mMCP-4) (*23*). In mMCP-$4^{-/-}$ and the respective wild-type control mice, intraperitoneal injection of S6b or *A.e.*v. induced comparable levels of mild hypothermia, intraperitoneal sarafotoxins, and PMC degranulation; moreover, all mice survived (fig. S6). Similar results were obtained when wild-type and mMCP-$4^{-/-}$ mice were injected intraperitoneally with ET-1 (fig. S7, A to C). These findings are in accord with the results of in vitro studies of rat (*22*) or mouse MCs (fig. S7, G and H) and our in vivo experiments in mice (fig. S7, D to F), all of which indicate that CPA is more effective than chymase in degrading ET-1.

We also used a lentivirus-based short hairpin–mediated RNA (shRNA) interference approach to generate mice with MC populations with normal or significantly decreased (>75%) levels of CPA activity (Fig. 2A). $Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$ mice that had been engrafted with either control (i.e., empty vector)–transduced MCs or CPA shRNA–transduced MCs exhibited similarities in numbers of PMCs (Fig. 2B), levels of PMC degranulation in response to S6b or *A.e.*v. (Fig. 2C), and appearance and distribution of MCs in the mesentery (Fig. 2D). However, upon intraperitoneal injection of S6b or *A.e.*v., CPA shRNA MC–engrafted $Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$ mice developed severe hypothermia, retained high concentrations of sarafotoxins in the peritoneal cavity, and died within 1 hour (Fig. 2, E and F), findings very similar to those observed in either MC-deficient $Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$ (Fig. 2, E and F) or $Kit^{W}/Kit^{W\text{-}v}$ (Fig. 1, A and B) mice, or in C57BL/6 wild-type mice that had been pretreated with PCI (Fig. 1, E and F).

To assess whether MCs could reduce the toxicity of snake venoms that do not contain



	Sarafotoxin (ng/peritoneum)	Survival
C57BL/6	0.02 ± 0.001	5 / 5
$Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$	85 ± 5.9 ***	0 / 7 ***
CPA shRNA MCs →$Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$	54 ± 9.7 ***	0 / 5 **
empty vector MCs →$Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$	0.1 ± 0.07	5 / 5

	Sarafotoxin (ng/peritoneum)	Survival
C57BL/6	0.02 ± 0.003	5 / 5
$Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$	75 ± 7.4 ***	0 / 5 **
CPA shRNA MCs →$Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$	62 ± 6 ***	0 / 5 **
empty vector MCs →$Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$	0.02 ± 0.001	5 / 5

Fig. 2. Reduction of CPA using shRNA in MCs renders mice susceptible to *A.e.*v.- and S6b-induced morbidity and mortality. (**A**) CPA activity in peritoneal cells from C57BL/6 or $Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$ mice engrafted with BMCMCs that were transduced with CPA shRNA or empty vector (control). $^*P < 0.05$ versus C57BL/6 or empty vector. (**B** to **D**) $Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$ mice engrafted with BMCMCs transduced with CPA shRNA or empty vector exhibit (B) normalization of PMC numbers, (C) similar extent of PMC degranulation in response to S6b and *A.e.*v., and (D) the appearance of MCs in the mesentery (arrows, MCs; scale bars, 100 μm). In (C), $^{***}P < 0.005$ versus C57BL/6; in (B) and (C), n.s., not significant ($P > 0.05$) versus values for C57BL/6 mice or for the comparisons indicated by brackets. (**E** and **F**) Changes in rectal temperature, levels of sarafotoxins in the peritoneal lavage fluid (at time of death or at 60 min, whichever was sooner), and 24-hour survival in C57BL/6 mice or $Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$ mice that were engrafted intraperitoneally with empty vector (empty vector MCs→$Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$) or CPA shRNA (CPA shRNA MCs→$Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$) expressing BMCMCs, and then were injected 4 weeks later with (E) 1 nmol S6b or (F) 10 μg *A.e.*v. $^{**}P < 0.01$ or $^{***}P < 0.005$ versus C57BL/6 or empty vector MCs→$Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$ mice.

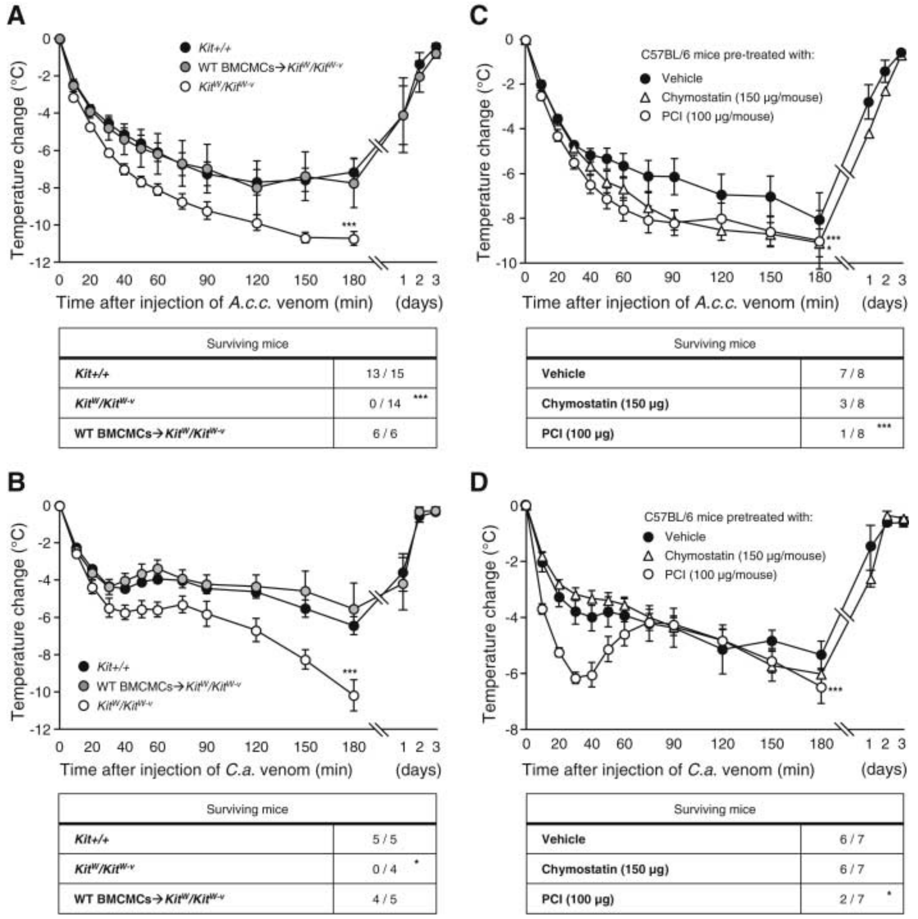
S6b or other ET-like peptides, we tested venoms from the western diamondback rattlesnake (*Crotalus atrox*, 150 μg) or the southern copperhead (*Agkistrodon contortrix contortrix*, 75 μg), species representative of pit vipers (Crotalidae), which account for the great majority of snake envenomations in North America (*24*). MC-deficient $Kit^{W}/Kit^{W\text{-}v}$ mice injected intraperitoneally with *Crotalus atrox* venom (*C.a*.v.) or *Agkistrodon contortrix contortrix* venom (*A.c.c*.v.) exhibited significantly lower body temperatures than wild-type mice, and all of them died within 24 hours. By contrast, most $Kit^{+/+}$ mice survived (87% in *C.a*.v.- and 100% in *A.c.c*.v.-injected mice) (Fig. 3, A and B), and these mice appeared to recover fully. Moreover, engraftment of $Kit^{W}/Kit^{W\text{-}v}$ mice with wild-type BMCMCs resulted in levels of protection against hypothermia and death that were statistically indistinguishable from those in wild-type mice (Fig. 3, A and B). Very similar results were obtained when these experiments were repeated with C57BL/6, $Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$, and wild-type MC-engrafted $Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$ mice (fig. S8).

We could not assess PMC degranulation in these experiments because of the excessive bleeding that had occurred within 60 min of intraperitoneal venom injection (or by the time of death) in all venom-injected mice. However, PMCs exhibited extensive degranulation when exposed to either type of venom ex vivo (fig. S9A) or, in C57BL/6 wild-type mice, at 10 min after intraperitoneal injection of venom in vivo (fig. S9, B and C).

MCs also diminished the extent of the hemorrhagic lesions that developed upon the injection of *C.a*.v. into the skin (fig. S10). Thus, MCs can diminish the local pathology induced by this venom, as well as reduce its systemic toxicity after intraperitoneal injection.

In wild-type (C57BL/6) mice injected intraperitoneally with venom, PCI pretreatment significantly worsened *C.a*.v.- or *A.c.c*.v.-induced hypothermia and mortality (Fig. 3, C and D) but did not result in levels of hypothermia or death rates (Fig. 3, C and D) that were quite as high as those observed in MC-deficient mice, especially in mice injected with *C.a*.v. One possible explanation for these results is that MC-derived proteases other than CPA (and/or additional MC-derived mediators or functions) may also have some protective effects against such venoms, and/or in counteracting the pathology they induce.

Like venomous reptiles, honeybees (*Apis mellifera*) and other stinging insects can account for substantial venom-associated morbidity and mortality (SOM Text 2). Although it is well-known that honeybee venom contains compounds that can induce MC degranulation (*25*), it had been thought that this contributed to the pathology associated with such stings (*26*). However, we found that MCs can confer significant protection against the hypothermia and death induced by the subcutaneous injection of *A. mellifera* venom (Fig. 4). Although all mice



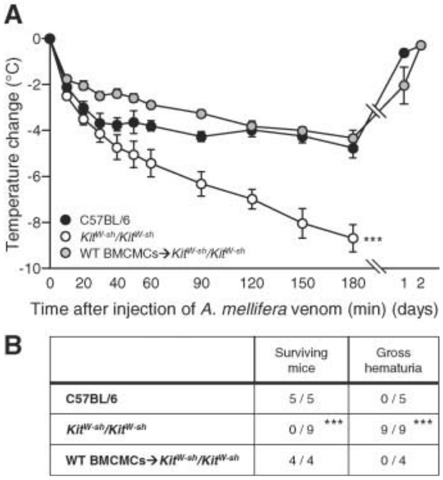
	Surviving mice
Kit+/+	13 / 15
KitW/KitW-v	0 / 14 ***
WT BMCMCs→KitW/KitW-v	6 / 6

	Surviving mice
Vehicle	7 / 8
Chymostatin (150 μg)	3 / 8
PCI (100 μg)	1 / 8 ***

	Surviving mice
Kit+/+	5 / 5
KitW/KitW-v	0 / 4 *
WT BMCMCs→KitW/KitW-v	4 / 5

	Surviving mice
Vehicle	6 / 7
Chymostatin (150 μg)	6 / 7
PCI (100 μg)	2 / 7 *

Fig. 3. MCs can limit the toxicity of western diamondback rattlesnake and southern copperhead venoms. Changes in rectal temperature and 24-hour survival after intraperitoneal injection of (**A** and **C**) *A.c.c.* venom (150 μg in 250 μl PBS) or (**B** and **D**) *C.a.* venom (75 μg in 250 μl) into (A and B) $Kit^{+/+}$ mice, MC-deficient $Kit^{W}/Kit^{W\text{-}v}$ mice, and $Kit^{W}/Kit^{W\text{-}v}$ mice that had been engrafted intraperitoneally with BMCMCs from $Kit^{+/+}$ mice (WT BMCMCs→$Kit^{W}/Kit^{W\text{-}v}$) or into (C and D) C57BL/6 mice that received intraperitoneal injections of vehicle (saline, 300 μl), chymostatin (150 μg in 300 μl of saline), or PCI (100 μg in 300 μl of saline) 10 min before intraperitoneal injection of snake venom. Mice that survived for 3 hours all appeared healthy and returned to baseline body temperature within 1 to 3 days. Body temperatures at 1 to 3 days were measured in surviving mice in one of the three experiments whose data were pooled to give the depicted 24-hour survival and up to 3 days of body temperature data. (A and B) *$P < 0.05$ or ***$P < 0.005$ versus $Kit^{+/+}$ or WT BMCMCs→$Kit^{W}/Kit^{W\text{-}v}$ mice; (C and D) *$P < 0.05$ or ***$P < 0.005$ versus vehicle.



	Surviving mice	Gross hematuria
C57BL/6	5 / 5	0 / 5
KitW-sh/KitW-sh	0 / 9 ***	9 / 9 ***
WT BMCMCs→KitW-sh/KitW-sh	4 / 4	0 / 4

Fig. 4. MCs can limit the toxicity of honeybee venom. (**A**) Changes in rectal temperature and (**B**) 24-hour survival and occurrence of gross hematuria after subcutaneous injection of *A. mellifera* venom (*A.m.*v.) at five different sites (three injections distributed over the length of the back skin and two into the belly skin, each containing 100 μg *A.m.*v. in 50 μl PBS) into WT mice, mast cell–deficient $Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$ mice, or $Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$ mice that had been engrafted intradermally, 6 weeks earlier, with 1.5 × 10^6 BMCMCs into each of the five injection areas (WT BMCMCs→$Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$). The amount of venom per injection (100 μg) roughly reflects the amount that can be delivered by one bee sting (*31*). All of the WT or WT BMCMCs→$Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$ mice appeared healthy, and their body temperatures returned to baseline within 2 days. ***$P < 0.005$ versus either C57BL/6 or WT BMCMCs→$Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$ mice.

exhibited the same initial response to such injections, namely, intense scratching of the injection sites, all of the MC-deficient $Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$ mice, but none of the wild-type mice or the wild-type BMCMC-engrafted $Kit^{W\text{-}sh}/Kit^{W\text{-}sh}$ mice, developed gross hematuria (Fig. 4).

Although the extent to which MCs might be able to enhance resistance to other animal venoms remains to be determined, components of the venoms of many different animals can activate MCs (*7*) (SOM Text 3). Moreover, it is quite possible that MC mediators, in addition to CPA and other proteases, also may contribute to the ability of MCs to reduce the morbidity and mortality associated with certain venoms. In 1965, Higginbotham hypothesized that MCs can reduce the toxicity of Russell's viper venom by degranulating and releasing heparin, which then binds highly cationic components of the venom and thereby reduces its toxicity (*27*). However, this interesting hypothesis has not yet been tested using MC-deficient and MC-engrafted mice (SOM Text 4). Finally, it is of course possible that MCs might contribute to (or have no effect on) the toxicity observed with some venoms.

We have identified a heretofore unproven role for MCs: enhancing innate host resistance to the toxicity of certain animal venoms. Our observations also provide a new perspective on the presence, within MCs, of prominent cytoplasmic granules that contain a large amount and, in some species, a large diversity, of proteases (*28*, *29*). It is likely that mast cell proteases can have beneficial roles in many settings, not only in host defense (*23*, *28*, *30*). However, we speculate that the storage in MC cytoplasmic granules of large amounts of proteases, which can be released to the exterior very rapidly upon suitable MC activation, reflects, at least in part, the selective pressure of the exposure of animals to diverse exogenous toxins (such as those in vertebrate and invertebrate venoms and perhaps those produced by certain microorganisms), as well as the advantage of being able to degrade and thereby control the toxicity of potent endogenous molecules such as ET-1 (*16*) (SOM Text 5).

References and Notes

1. B. G. Fry *et al.*, *Nature* **439**, 584 (2006).
2. J. P. Chippaux, *Bull. World Health Organ.* **76**, 515 (1998).
3. R. D. Theakston, D. A. Warrell, E. Griffiths, *Toxicon* **41**, 541 (2003).
4. A. Rucavado, T. Escalante, J. M. Gutierrez, *Toxicon* **43**, 417 (2004).
5. J. White, *Toxicon* **45**, 951 (2005).
6. N. K. Dutta, K. G. Narayanan, *Nature* **169**, 1064 (1952).
7. A. Weisel-Eichler, F. Libersat, *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **190**, 683 (2004).
8. D. D. Metcalfe, D. Baram, Y. A. Mekori, *Physiol. Rev.* **77**, 1033 (1997).
9. M. F. Gurish, K. F. Austen, *J. Exp. Med.* **194**, F1 (2001).
10. H. Turner, J. P. Kinet, *Nature* **402**, B24 (1999).
11. F. D. Finkelman, M. E. Rothenberg, E. B. Brandt, S. C. Morris, R. T. Strait, *J. Allergy Clin. Immunol.* **115**, 449 (2005).
12. B. Echtenacher, D. N. Männel, L. Hültner, *Nature* **381**, 75 (1996).
13. R. Malaviya, T. Ikeda, E. Ross, S. N. Abraham, *Nature* **381**, 77 (1996).
14. A. P. Prodeus, X. Zhou, M. Maurer, S. J. Galli, M. C. Carroll, *Nature* **390**, 172 (1997).
15. M. Maurer *et al.*, *J. Exp. Med.* **188**, 2343 (1998).
16. M. Maurer *et al.*, *Nature* **432**, 512 (2004).
17. Y. Kloog *et al.*, *Science* **242**, 268 (1988).
18. Materials and methods are available as supporting material on *Science* Online.
19. E. Kochva, *Public Health Rev.* **26**, 209 (1998).
20. T. Nakano *et al.*, *J. Exp. Med.* **162**, 1025 (1985).
21. M. Tsai *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9186 (2000).
22. K. P. Metsärinne *et al.*, *Arterioscler. Thromb. Vasc. Biol.* **22**, 268 (2002).
23. E. Tchougounova, G. Pejler, M. Abrink, *J. Exp. Med.* **198**, 423 (2003).
24. D. McNamee, *Lancet* **357**, 1680 (2001).
25. M. R. Ziai, S. Russek, H. C. Wang, B. Beer, A. J. Blume, *J. Pharm. Pharmacol.* **42**, 457 (1990).
26. M. C. Calixto, K. M. Triches, J. B. Calixto, *Inflamm. Res.* **52**, 132 (2003).
27. R. D. Higginbotham, *J. Immunol.* **95**, 867 (1965).
28. C. Huang, A. Sali, R. L. Stevens, *J. Clin. Immunol.* **18**, 169 (1998).
29. G. H. Caughey, *Mol. Immunol.* **38**, 1353 (2002).
30. J. Mallen-St. Clair, C. T. Pham, S. A. Villalta, G. H. Caughey, P. J. Wolters, *J. Clin. Invest.* **113**, 628 (2004).
31. J. O. Schmidt, *Toxicon* **33**, 917 (1995).

32. We thank E. Kochva and A. Bdolah for providing *Atractaspis engaddensis* venom and for helpful discussions, D. E. Clouthier and M. M. Yanagisawa for providing $ET_A^{-/-}$ and $ET_A^{+/-}$ ES cells, R. L. Stevens and R. Adachi for calling our attention to the similarities between vertebrate mast cells and ascidian test cells, and M. Liebersbach and A. Xu for technical assistance. This work was supported by NIH grants R37 AI23990, R01 CA72074, and P50 HL67674 (to S.J.G.) and by grants from the Deutsche Forschungsgemeinschaft (ME2668/1-1 to M.M.) and the Boehringer Ingelheim Fonds (to V.L.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/313/5786/526/DC1
Materials and Methods
SOM Text
Figs. S1 to S10
References

18 April 2006; accepted 15 June 2006
10.1126/science.1128877

Multiple Phosphorylation Sites Confer Reproducibility of the Rod's Single-Photon Responses

Thuy Doan,[1] Ana Mendez,[4] Peter B. Detwiler,[2] Jeannie Chen,[4]* Fred Rieke[2,3]*

Although signals controlled by single molecules are expected to be inherently variable, rod photoreceptors generate reproducible responses to single absorbed photons. We show that this unexpected reproducibility—the consistency of amplitude and duration of rhodopsin activity—varies in a graded and systematic manner with the number but not the identity of phosphorylation sites on rhodopsin's C terminus. These results indicate that each phosphorylation site provides an independent step in rhodopsin deactivation and that collectively these steps tightly control rhodopsin's active lifetime. Other G protein cascades may exploit a similar mechanism to encode accurately the timing and number of receptor activation.

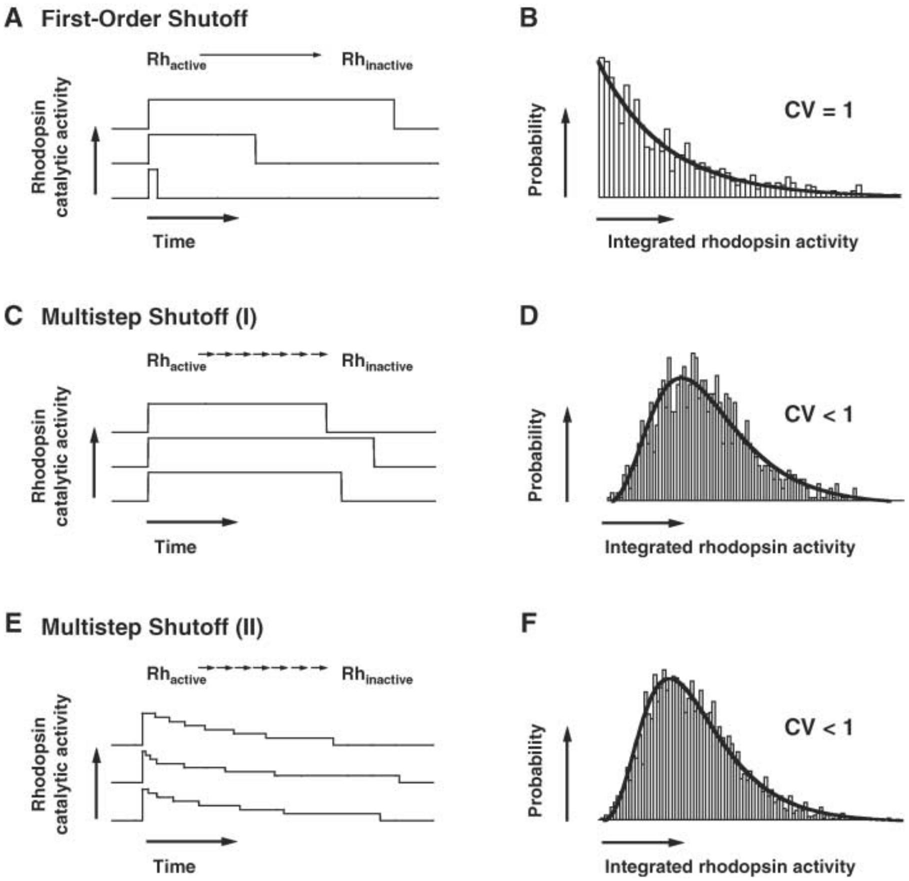
Rhodopsin may be biology's most precise single-molecule timekeeper. In retinal rod photoreceptors, the effective absorption of a single photon activates a single rhodopsin molecule, which triggers a highly amplified signal transduction cascade to produce a macroscopic change in the current flowing into the outer segment of the receptor. The electrical response evoked by a single photon shows much less trial-to-trial variability than other familiar signals generated by single molecules, such as the time to decay of a radioactive particle or the charge flowing through an ion channel during a single opening (*1*–*5*). More generally, events controlled by a first-order process (Fig. 1, A and B) are inherently more variable than the responses to single photons. Previous studies indicate that the low variability in the rod's current responses arises from low variability in the lifetime of light-activated rhodopsin (*2*–*6*). This poses a challenging molecular design problem: How is the active lifetime of a single molecule regulated so tightly?

Past work argues, largely by excluding other explanations, that reproducibility is produced by the deactivation, or shutoff, of a single rhodopsin molecule through a series of steps or transitions (*2*–*6*). The essence of this multistep shutoff model is simple averaging: The integrated rhodopsin activity averaged over multiple stochastic steps varies less than the activity controlled by a single step. Rhodopsin activity could be timed by the occurrence of each step before terminating with completion of the final step (Fig. 1, C and D), or rhodopsin activity could decline after each step (Fig. 1, E and F). In either case, the coefficient of variation (CV = standard deviation/mean) for rhodopsin's integrated activity is $1/\sqrt{N}$ for N independent steps that each control an equal frac-

[1]Program in Neurobiology and Behavior, [2]Department of Physiology and Biophysics, [3]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA. [4]Department of Ophthalmology and Cell and Neurobiology Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA.

*To whom correspondence should be addressed. E-mail: rieke@u.washington.edu (F.R.); jeannie@usc.edu (J.C.)

Fig. 1. Proposed models of rhodopsin shutoff. (**A**) Simulated activity of rhodopsin after a single stochastic shutoff step, (**C**) seven independent steps of equal rate constant where only the final step affects rhodopsin activity, and (**E**) seven independent steps where each step reduces an equal fraction of rhodopsin activity. (**B**, **D**, and **F**) Distributions of integrated rhodopsin activity from 1000 simulated responses, as in (A), (C), and (E). A first-order shutoff model predicts that the integrated rhodopsin activity is exponentially distributed with a $CV = 1$, whereas either multistep shutoff model predicts a $CV < 1$.



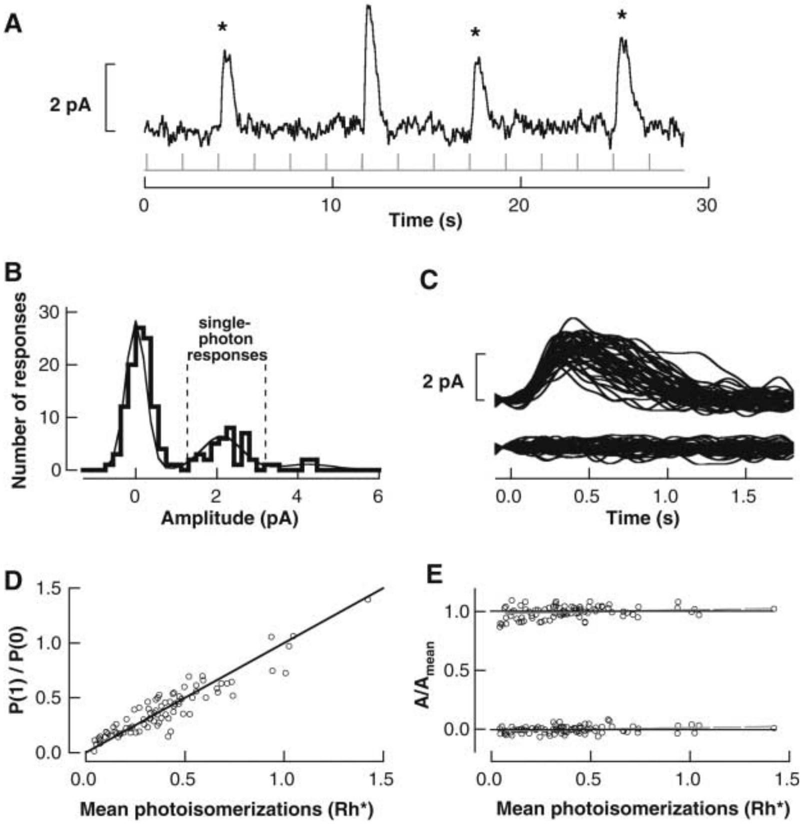


Fig. 2. Identification of single-photon responses in wild-type mouse rods. (**A**) Current responses recorded at 30°C to fixed-strength flashes (vertical bars) that produced on average 0.4 isomerizations. Asterisks indicate identified single-photon responses. Dark current = 17 pA; bandwith = 0 to 5 Hz. (**B**) Histogram of response amplitudes from the same rod and flash strength as (A). The fit (thin curve) was calculated using Eq. 1 [available in (*15*)]. Vertical dashed lines represent thresholds used to identify single-photon responses (*15*). (**C**) Graph of 50 consecutive single-photon responses and responses to zero absorbed photons isolated from the same rod as (A). (**D**) Average Rh* estimated from the ratio of the number of identified single-photon responses [$P(1)$] and responses to zero absorbed photons [$P(0)$] plotted against the average Rh* estimated from the collecting area and flash strength (29 cells). The points fall near the line of unity slope, indicating that the number of identified single-photon responses and responses to zero photons were consistent with expectations from Poisson statistics. Thus, systematic biases in the identification procedure were small. (**E**) Amplitudes of single-photon responses and responses to zero absorbed photons plotted against the strength of the flash (in Rh*) from which responses were identified. For each cell ($n = 29$), amplitudes (A) were normalized by the mean single-photon response amplitude (A_{mean}) for all flash strengths. Error-free identification predicts no dependence of the amplitude of the mean single-photon responses and responses to zero absorbed photons on flash strength (black lines). Slopes of the best-fit lines (dashed gray) through the data were 0.030 ± 0.003 (mean ± SD) for single-photon responses and 0.026 ± 0.002 for responses to zero absorbed photons, indicating that >90% of the single-photon responses were identified correctly. A similar low probability of errors in the identification of single-photon responses held for wild-type and transgenic rods.

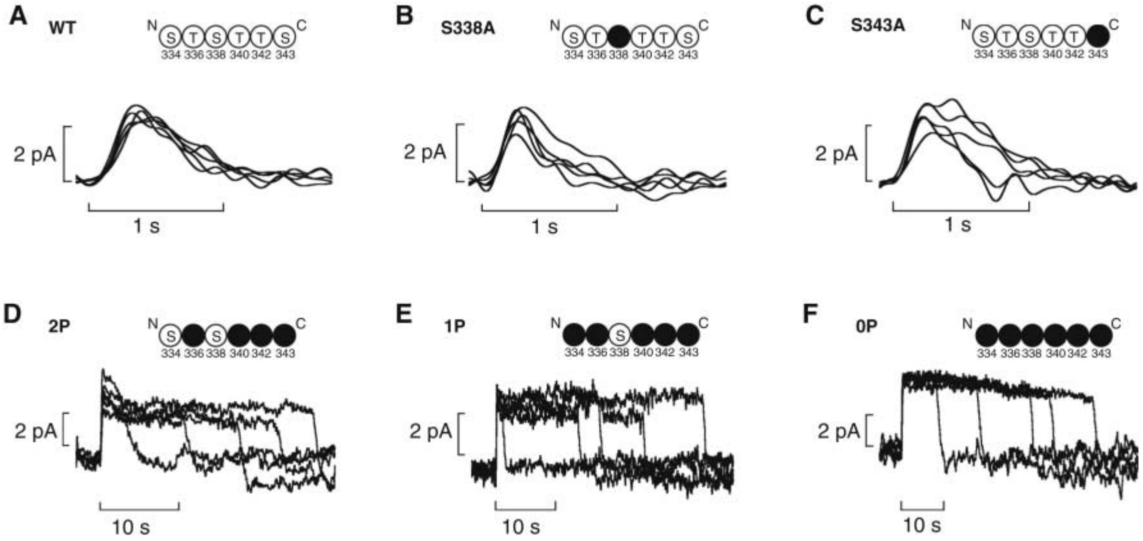


Fig. 3. Examples of single-photon responses produced by wild-type and mutated rhodopsin. Five identified single-photon responses from a mouse rod expressing (**A**) wild-type (WT) rhodopsin, (**B** and **C**) rhodopsin with five phosphorylation sites (S338A and S343A), (**D**) rhodopsin with two sites (2P), (**E**) rhodopsin with one site (1P), and (**F**) rhodopsin with zero sites (0P). Insets show simplified schematics of the C-terminal residues. Only the potential serine (S)/threonine (T) phosphorylation sites are shown. Black circles represent sites mutated to alanine. Response variability increases as the number of remaining phosphorylation sites decreases.

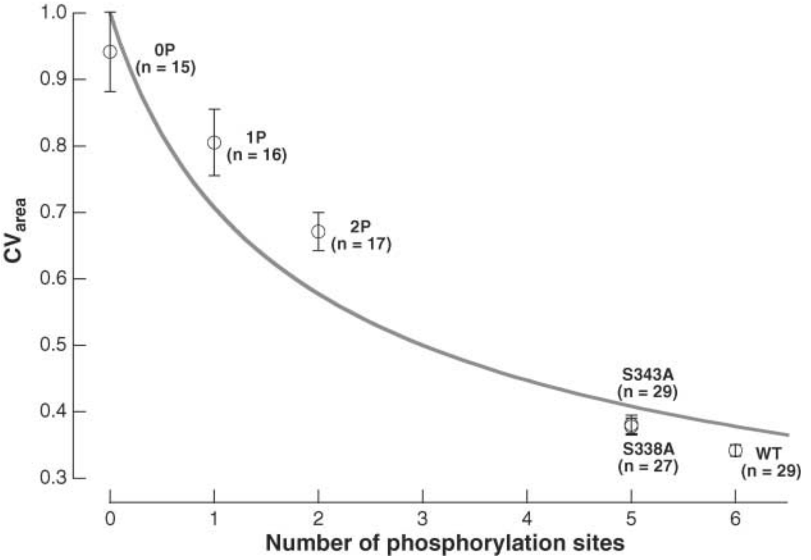


Fig. 4. Correlation of single-photon response variability with number of rhodopsin phosphorylation sites. Circles and vertical bars plot the mean CV_{area} ± SEM. The smooth curve is the CV_{area} predicted by $1/\sqrt{N_{\text{P}} + 1}$, where N_{P} is the number of phosphorylation sites and 1 represents arrestin binding.

tion of rhodopsin's integrated catalytic activity. Rhodopsin activity, similar to that of most G protein–coupled receptors, is terminated by phosphorylation and arrestin binding (*7–12*), leading to the proposal that multiple phosphorylations of rhodopsin provide the molecular steps for the multistep shutoff model (*4*, *13*, *14*). This hypothesis, however, has not been tested directly. The key missing element has been the ability to identify single-photon responses and quantify the effect of genetic manipulations on the activity of single rhodopsin molecules.

We characterized the variability of identified single-photon responses of wild-type and transgenic mouse rods to determine whether reproducibility depends on the number of rhodopsin phosphorylation sites in the graded and systematic manner predicted by the multistep shutoff model. From single-cell recordings of the current flowing into the rod outer segment, we identified single-photon responses among responses to dim flashes producing on average 0.07 to 1.5 absorbed photons (Rh*) (*15*). These measured currents reflected the activity of light-activated rhodopsin molecules. Clear identification of single-photon responses requires the electrical current in response to the absorption of a photon to be distinguishable from the background current fluctuations and responses to multiple photons, requirements that were met in the recordings used here (Fig. 2, A and C) (*15*). Single-photon responses were separated from responses to zero and multiple absorbed photons by applying thresholds to histograms of the amplitude of the responses to a repeated dim flash (Fig. 2B). Tests for errors in identification (Fig. 2, D and E) indicated that <10% of the true single-photon responses failed to be identified and <10% of the identified responses were to zero or multiple absorbed photons (*5*, *15*).

We used the *CV* of the response areas (CV_{area}, the time integral of the response) to characterize the variability of the identified single-photon responses. The CV_{area} captures the total variability of the response, independent of whether the variability occurs early or late (fig. S1) (*5*, *14*). In the absence of saturation within the transduction cascade, the CV_{area} provides an upper bound on the variability of rhodopsin activity because components downstream of rhodopsin could also contribute. The CV_{area} for wild-type single-photon responses was 0.34 ± 0.01 at 30°C (mean ± SEM, $n = 29$) and 0.36 ± 0.02 at 35°C ($n = 27$). We performed all subsequent experiments at 30°C because the identification of single-photon responses was cleaner. Errors in identification of single-photon responses did not substantially influence the measured variability (*15*).

The proposal that the steps in the multistep shutoff model are provided by phosphorylation predicts that the CV_{area} of the single-photon responses will increase as the number of phosphorylation sites decreases, scaling as $1/\sqrt{N}$ for *N* independent and equal steps. We tested this prediction by quantifying the variability of identified single-photon responses in mouse rods expressing rhodopsin mutants with five, two, one, and zero remaining phosphorylation sites; wild-type rhodopsin has six phosphorylation sites (*4*). The number of sites was reduced by mutating serine or threonine residues to alanine (Fig. 3, insets) (*4*). Qualitatively, variability increased as the number of steps decreased (Fig. 3). Quantitatively, the CV_{area} in each case was near the $1/\sqrt{N}$ prediction, assuming that arrestin binding provided a final step in quenching rhodopsin activity (Fig. 4) (*10*).

The multistep shutoff model further predicts that reducing the number of phosphorylation sites from six to five should produce an ~8% increase in CV_{area}, independent of which site is removed. To test this prediction, we characterized single-photon responses from two rhodopsin mutants [$\text{Ser}^{338}\rightarrow\text{Ala}^{338}$ (S338A) and $\text{Ser}^{343}\rightarrow\text{Ala}^{343}$ (S343A)] with five phosphorylation sites (*4*). The measured CV_{area} was 0.38 ± 0.01 for S338A (n = 27) and 0.38 ± 0.02 for S343A (n = 29); both were significantly greater than the CV_{area} of 0.34 ± 0.01 of wild-type rods ($P < 0.001$ for S338A and $P < 0.05$ for S343A, Student's *t* test). The increase in CV_{area} was ~10%, very close to that predicted from the multistep shutoff model. The similar increase in variability of the S338A and S343A single-photon responses indicates that these sites make equal contributions to regulating rhodopsin shutoff.

The long duration of the single-photon responses in the mutants with one and two remaining phosphorylation sites suggested that a single shutoff step limited the duration of rhodopsin activity. If that is correct, rhodopsin shutoff should effectively be a first-order process, with a CV_{area} of 1 and an exponential distribution of lifetimes (Fig. 1, A and B). Thus, we were surprised to find a $CV_{\text{area}} < 1$ in these mutants. Whereas the durations of the single-photon responses produced by rhodopsin with zero phosphorylation sites were exponentially distributed (fig. S2C), those produced by rhodopsin with one or two remaining sites were not (fig. S2, A and B; $P < 0.05$ for two sites; $P < 0.002$ for one site, χ^2 test). The nonexponential lifetime distribution and CV_{area} less than 1 indicate either that phosphorylation occurs slowly in these mutants, and hence rhodopsin shutoff is still effectively controlled by several steps, or that arrestin binding itself is less stochastic than expected from a first-order process (*16*).

Although phosphorylation and arrestin binding are common steps in the shutoff of G protein–coupled receptors, many questions remain about the molecular details of these events. Our results provide some constraints. First, arrestin binding apparently controls a small fraction of the integrated rhodopsin activity under normal conditions, requiring that arrestin rarely binds to rhodopsin until most or all phosphorylation events have been completed. Indeed, biochemical measurements indicate that the affinity of rhodopsin for arrestin binding to rhodopsin increases as the number of phosphorylations increases (*13*, *17*). Second, each phosphorylation site could make an equal contribution to the shutoff of rhodopsin in one of two ways: (i) Rhodopsin's catalytic activity remains constant until arrestin binds and the rate constants associated with each phosphorylation event are equal (Fig. 1C), or (ii) each phosphorylation event decreases rhodopsin's catalytic activity and slows subsequent phosphorylation events (Fig. 1E). Previous measurements support the latter model (*5*, *10*, *13*), although combinations of the two are also possible.

We found that variability of the single-photon responses depended on the number of phosphorylation sites in a systematic and graded manner, including an increase in variability when a single site was removed. This result differs from past work that found that variability remained constant or decreased when one or three phosphorylation sites were removed (*4*). These previous experiments, however, were based on indirect estimates of single-photon response variability and thus could have missed the subtle increase in variability associated with the removal of a few phosphorylation sites. Although our results show that multistep shutoff through phosphorylation and arrestin binding is the dominant factor limiting variability of the rod's single-photon responses, we cannot rule out smaller contributions from other mechanisms. Indeed, single-photon responses of primate rods vary 20 to 30% less than those of mouse rods despite identical numbers of phosphorylation sites, suggesting that other mechanisms such as local saturation of the transduction cascade may help reduce response variability (*5*, *18*).

Vision in starlight requires detecting and processing responses to individual photons. Under these conditions, reproducibility of the single-photon response permits rods to encode accurate information about the number and timing of photon absorptions. Other G protein cascades—e.g., those in pheromone receptors (*19*, *20*)—operate when few receptors are active. Thus, receptor shutoff through multiple steps may be a general strategy to improve the fidelity of signals generated by G protein cascades.

References and Notes

1. D. Baylor, T. Lamb, K. Yau, *J. Physiol.* **288**, 613 (1979).
2. F. Rieke, D. Baylor, *Biophys. J.* **75**, 1836 (1998).
3. G. Whitlock, T. Lamb, *Neuron* **23**, 337 (1999).
4. A. Mendez *et al.*, *Neuron* **28**, 153 (2000).
5. G. Field, F. Rieke, *Neuron* **35**, 733 (2002).
6. M. Burns, A. Mendez, J. Chen, D. Baylor, *Neuron* **36**, 81 (2002).
7. U. Wilden, H. Kuhn, *Biochemistry* **21**, 3014 (1982).
8. H. Kuhn, S. Hall, U. Wilden, *FEBS Lett.* **176**, 473 (1984).
9. J. Chen, C. Makino, N. Peachey, D. Baylor, M. Simon, *Science* **267**, 374 (1995).
10. J. Xu *et al.*, *Nature* **389**, 505 (1997).
11. C. Chen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3718 (1999).
12. M. Kennedy *et al.*, *Neuron* **31**, 87 (2001).
13. S. Gibson, J. Parkes, P. Liebman, *Biochemistry* **39**, 5738 (2000).
14. R. Hamer, S. Nicholas, D. Tranchina, P. Liebman, T. Lamb, *J. Gen. Physiol.* **122**, 419 (2003).
15. Materials and methods are available as supporting material on *Science* Online.
16. V. Gurevich, E. Gurevich, *Trends Pharmacol. Sci.* **25**, 105 (2004).
17. U. Wilden, *Biochemistry* **34**, 1446 (1995).
18. S. Ramanathan, P. Detwiler, A. Sengupta, B. Shraiman, *Biophys. J.* **88**, 3063 (2005).
19. T. Leinders-Zufall *et al.*, *Nature* **405**, 792 (2000).
20. F. Zufall, K. Kelliher, T. Leinders-Zufall, *Microsc. Res. Tech.* **58**, 251 (2002).
21. We thank B. Hille, G. Murphy, F. Dunn, C. Asbury, B. Pinsky, J. Jensen, B. Wark, and A. Hinterwirth for constructive comments on the manuscript. Support was provided by the NIH through grants EY-11850 (F.R.), EY-12155 (J.C.), EY-02048 (P.B.D.), and T32EY-07031 (T.D.); Poncin Scholarship (T.D.); and the Howard Hughes Medical Institute (F.R.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/313/5786/530/DC1
Materials and Methods
Figs. S1 and S2
References

23 February 2006; accepted 6 June 2006
10.1126/science.1126612

Activated Signal Transduction Kinases Frequently Occupy Target Genes

Dmitry K. Pokholok,[1]* Julia Zeitlinger,[1]* Nancy M. Hannett,[1] David B. Reynolds,[1] Richard A. Young[1,2]†

Cellular signal transduction pathways modify gene expression programs in response to changes in the environment, but the mechanisms by which these pathways regulate populations of genes under their control are not entirely understood. We present evidence that most mitogen-activated protein kinases and protein kinase A subunits become physically associated with the genes that they regulate in the yeast (*Saccharomyces cerevisiae*) genome. The ability to detect this interaction of signaling kinases with target genes can be used to more precisely and comprehensively map the regulatory circuitry that eukaryotic cells use to respond to their environment.

Signal transduction pathways mediate the cellular response to specific environmental or developmental signals. The activation of signal transduction pathways can lead to phosphorylation of transcription factors (*1*, *2*), histones (*3*), chromatin-modifying complexes, and the transcription machinery (*4*, *5*). These modifications contribute to changes in the gene expression program. Although the traditional view has been that most phosphorylation events do not occur at the genes that are ultimately controlled by signal transduction pathways, recent reports have revealed that at least one mitogen-activated protein kinase (MAPK)—high-osmolarity glycerol 1p (Hog1p) in yeast and its homolog p38 in

[1] Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142, USA. [2] Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: young@wi.mit.edu

humans—physically occupies certain genes where it regulates gene expression [(6–10), reviewed in (5)]. This evidence, and the knowledge that the terminal kinases of multiple signal transduction pathways can be found in the nucleus under activating conditions (*11*), led us to investigate the possibility that components of activated signal transduction pathways generally become associated with chromatin at the genes that they activate.

To confirm previous reports that the MAPK Hog1p in yeast occupies genes upon exposure of cells to osmotic stress and to identify the complete set of genes that were so occupied, we performed chromatin immunoprecipitation coupled with microarrays (ChIP-Chip) experiments using a yeast strain in which endogenous Hog1p has a tandem affinity purification (TAP) tag (Fig. 1). The presence of the TAP tag was verified for this and all other yeast strains used in this study (fig. S1). In cells exposed to 0.4 M NaCl for 5 min (*6*), we identified 36 genes that were occupied by Hog1p at high confidence and showed increased transcription after exposure to NaCl or KCl (Fig. 1B). Among these were all of the seven genes previously found to be occupied by Hog1p (*6*, *7*, *12*) (Fig. 1B). We detected little occupancy of Hog1p at genes before osmotic stress, which is consistent with evidence that Hog1p is translocated into the nucleus during osmotic stress (*11*, *13*, *14*). Most genes that were occupied by Hog1p during osmotic stress showed an altered expression pattern in cells lacking Hog1p (Fig. 1B). Hog1p occupancy was highest at the promoters of genes but was also observed throughout the entire transcribed region of these genes (Fig. 1C and fig. S2).

The MAPKs Fus3p and Kss1p are activated in response to pheromone exposure and induce the expression of mating genes in yeast (*15*) (Fig. 2A). We used genome-wide ChIP-Chip analysis to determine whether Fus3p and Kss1p occupy a specific set of genes upon activation (Fig. 2). Nine genes were occupied by Fus3p and showed increased transcription within 5 min after exposure to mating pheromone (Fig. 2B). Essentially the same set of genes was occupied by Kss1p (Fig. 2D). These genes were previously shown to be dependent on the pheromone MAPK pathway for their expression (*16*). Enrichment of both kinases was observed throughout the transcribed regions of these genes (Fig. 2, C and E).

Ste5p, the central scaffold protein of the pheromone response pathway, interacts with Fus3p and possibly Kss1p at the plasma membrane but can also be found in the nucleus (*11*, *15*, *17*). We found that TAP-tagged Ste5p occupied essentially the same mating genes that were bound by Fus3p and Kss1p (Fig. 2F). Ste5p was observed throughout the transcribed regions of these genes (Fig. 2G). These results suggest that Ste5p may function as an adaptor for protein-protein interactions both at the plasma membrane and in the nucleus.

The cyclic adenosine monophosphate (cAMP)–activated protein kinase A (PKA) is stimulated by an increased concentration of intracellular cAMP when yeast are exposed to fermentable carbon sources such as glucose (Fig. 3A) (*18*). There are three PKA catalytic subunits in yeast: Tpk1p, Tpk2p, and Tpk3p. Genome-wide ChIP-Chip analyses suggested that Tpk1p occupies the entire transcribed region of most actively transcribed genes in cells grown in glucose media (Fig. 3, B and C). To further test this possibility, we determined whether Tpk1p occupancy would change at genes that were dynamically repressed or activated as yeast cells were subjected to different environmental conditions. Indeed, Tpk1p occupancy was reduced at genes whose expression was reduced when cells were transferred to a nonfermentable carbon source (glycerol) (Fig. 3D). In contrast, Tpk1p became associated with genes that were activated when

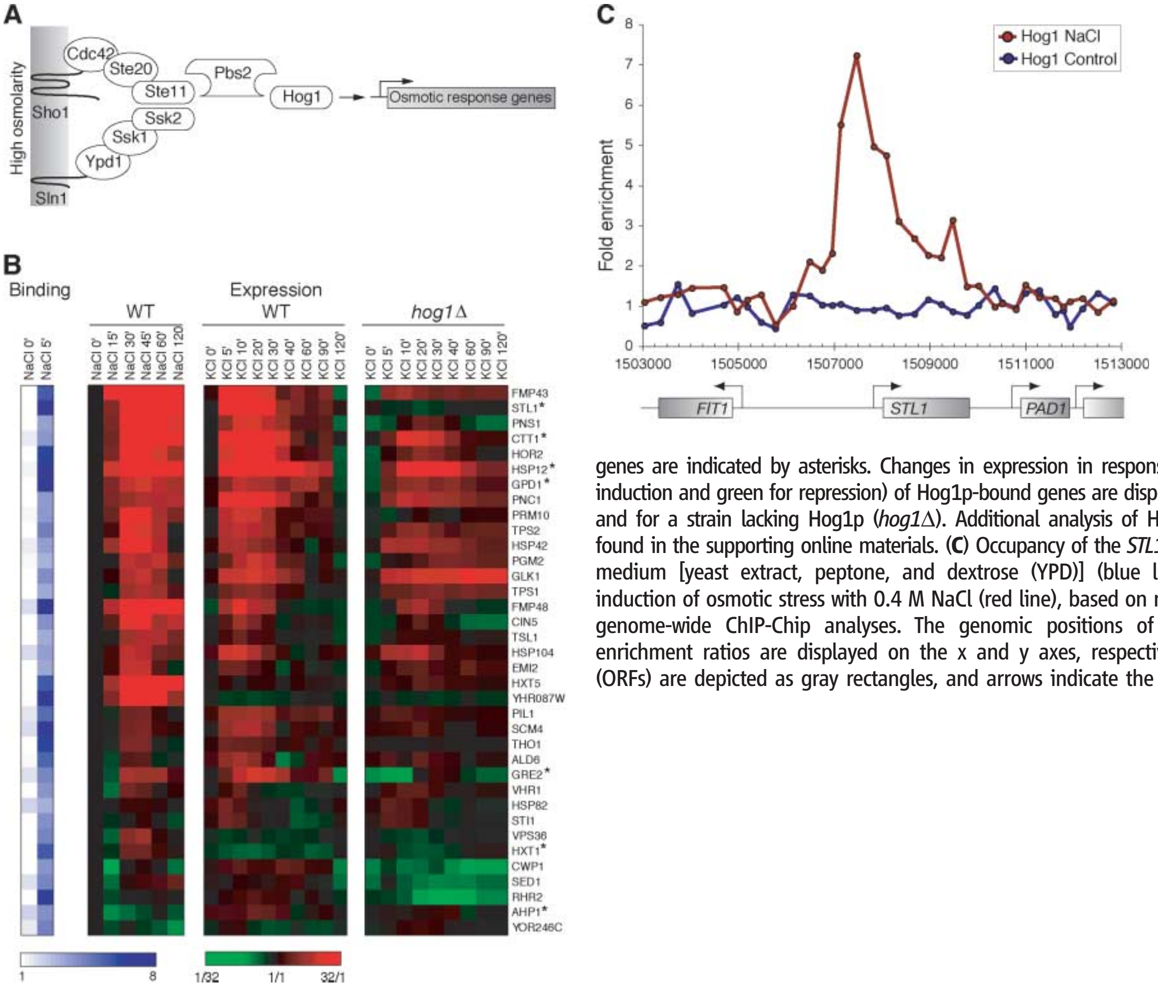


Fig. 1. Recruitment of Hog1p to promoters and transcribed regions of genes activated by osmotic stress. (**A**) The Hog MAPK pathway in *S. cerevisiae*. (**B**) Genes that are bound by Hog1p (table S2) and induced during osmotic stress (with NaCl or KCl). The maximum ChIP enrichment of Hog1p along each gene before and 5 min after NaCl addition are shown in blue. Previously identified target genes are indicated by asterisks. Changes in expression in response to osmotic stress (red for induction and green for repression) of Hog1p-bound genes are displayed for wild-type cells (WT) and for a strain lacking Hog1p (*hog1*Δ). Additional analysis of Hog1-occupied genes can be found in the supporting online materials. (**C**) Occupancy of the *STL1* gene by Hog1p in control medium [yeast extract, peptone, and dextrose (YPD)] (blue line) and 5 min after the induction of osmotic stress with 0.4 M NaCl (red line), based on merged duplicate data from genome-wide ChIP-Chip analyses. The genomic positions of probe regions and their enrichment ratios are displayed on the x and y axes, respectively. Open reading frames (ORFs) are depicted as gray rectangles, and arrows indicate the direction of transcription.

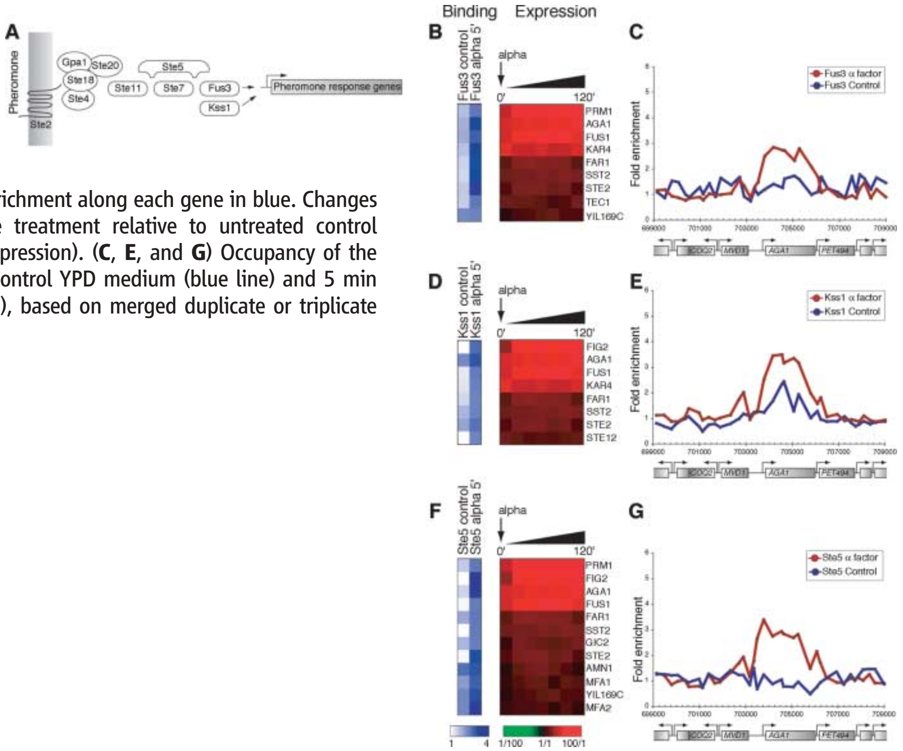


Fig. 2. Recruitment of Fus3p, Kss1p, and Ste5p to genes expressed in response to alpha pheromone treatment. (**A**) The pheromone response MAPK pathway in *S. cerevisiae*. (**B**, **D**, and **F**) Genes bound by Fus3p (B), Kss1p (D), and Ste5p (F) at high confidence (tables S3 to S5) that are induced after alpha pheromone treatment. The occupancy of Fus3p, Kss1p, and Ste5p is shown as maximum ChIP enrichment along each gene in blue. Changes in the expression of these genes during pheromone treatment relative to untreated control samples are displayed in red (induction) and green (repression). (**C**, **E**, and **G**) Occupancy of the *AGA1* gene by Fus3p (C), Kss1p (E), and Ste5p (G) in control YPD medium (blue line) and 5 min after exposure to alpha pheromone (5 μg/ml) (red line), based on merged duplicate or triplicate data from genome-wide ChIP-Chip analyses.

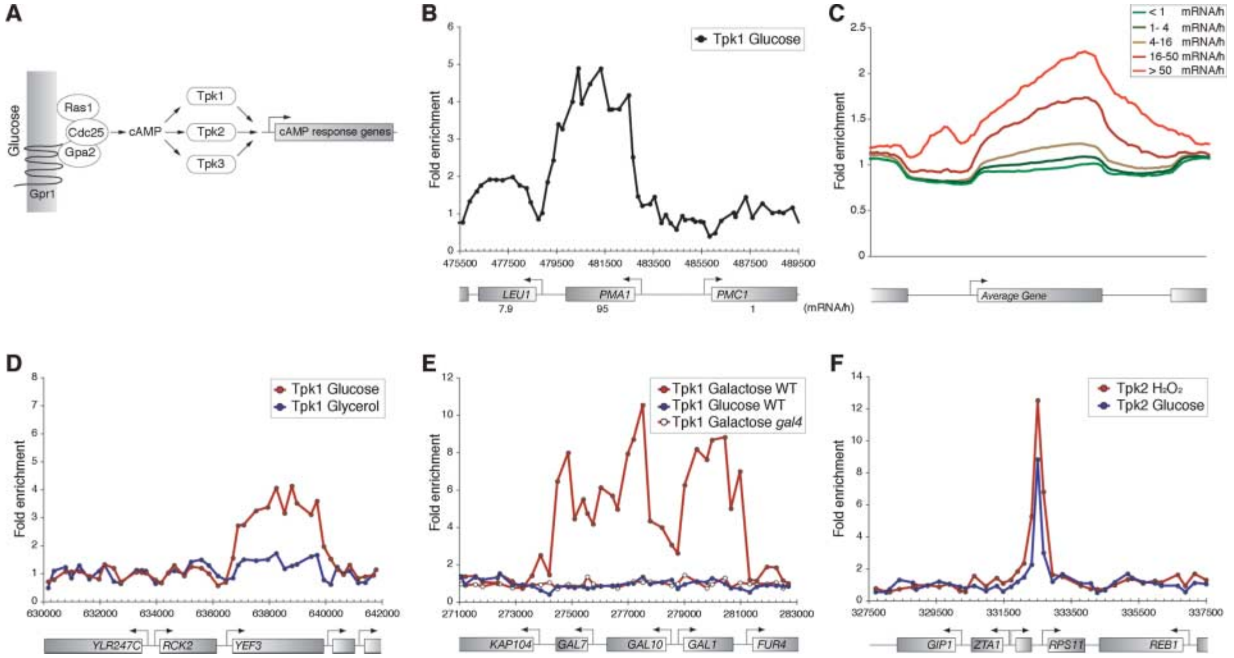


Fig. 3. Occupancy of transcribed regions of active genes by Tpk1p and of promoters of ribosomal protein genes by Tpk2p. (**A**) The cAMP/PKA signaling pathway in *S. cerevisiae*. (**B**) Occupancy of Tpk1p at a portion of chromosome VII, containing the *PMA1* and *LEU1* genes, in the presence of glucose based on data from genome-wide ChIP-Chip analyses. The transcriptional frequency of the corresponding ORF (*26*) is indicated as mRNA per hour underneath each ORF. (**C**) Average Tpk1p enrichment for classes of different transcriptional frequencies [determined by means of metagene analysis (*27*)]. The genome's 5324 genes were divided into five classes according to their transcriptional rate (*26*). A fixed length was assigned to each ORF and intergenic region, and probes were assigned to the nearest relative position and averaged for each class. (**D**) Tpk1p occupancy at a portion of chromosome XII, containing the *YEF3* gene whose transcription is substantially reduced during growth in medium containing glycerol (blue line) as compared to that in control medium (YPD) containing glucose (red line). (**E**) Tpk1p occupancy at *GAL1-10* locus in glucose-containing medium (blue line) after the addition of galactose (red line with solid circles) and of galactose in the absence of Gal4p (*gal4*Δ) (red line with open circles). (**F**) Tpk2p occupancy at the promoter of the *RPS11B* gene during oxidative stress and in control medium (YPD) containing glucose.

cells were exposed to galactose (Fig. 3E). Occupancy at these galactose-inducible genes was dependent on gene activation because it was not detected in strains lacking the transcriptional activator Gal4p (Fig. 3E). These results confirm that Tpk1p generally becomes physically associated with actively transcribed genes and that occupancy occurs throughout the transcribed portions of these genes.

We then investigated whether Tpk2p occupies specific portions of the genome. Tpk2p was found almost exclusively associated with the promoters of ribosomal protein genes (Fig. 3F, fig. S3, and table S6). Gene occupancy by Tpk2p did not correlate with transcription rates throughout the genome, and Tpk2p remained associated with its target genes when cells were exposed to oxidative stress, which leads to reduced transcription of ribosomal protein genes (Fig. 3F). We did not detect Tpk3p occupancy on chromatin under the conditions used here (rich media, oxidative stress, and pheromone exposure). Although we have not shown that occupancy of genes by Tpk1p and Tpk2p regulates gene expression, previous studies have shown that PKA phosphorylates the Srb9 subunit of the Mediator complex (*19*) and that PKA activity regulates ribosomal gene expression (*20–22*). The idea that some PKA family members might operate, at least in part, through occupancy of actively transcribed genes is attractive because it might provide an efficient means for cells to respond to the nutrient environment at the level of gene expression.

Our finding that most activated MAPKs and PKAs in yeast become associated with distinct target genes changes our perception of the sites at which signaling pathways act to regulate gene expression. With the exception of Hog1p and p38, studies of the effect of signal transduction pathways on gene expression have not implied that the activities of MAPKs or PKAs involve genome occupancy. Although it is still possible that the phosphorylation of transcriptional regulators also occurs elsewhere in the cell, the detection of kinases by ChIP-Chip analyses at target genes suggests a model in which regulation by signal transduction kinases often occurs at the genes themselves. In this model, kinases become physically localized at specific sites in the genome by association with transcription factors, chromatin regulators, the transcription apparatus, nucleosomes, or nuclear pore proteins that are associated with subsets of actively transcribed genes (*5–10*, *19*, *23–25*) (fig. S4).

The kinases studied here associate with target genes in at least three different patterns, suggesting that there are different mechanisms involved in their association with genes. Tpk2p was found only at the promoter regions of its target genes. Hog1p occupancy was greatest at the promoters but also occurred to a limited extent within the transcribed regions of genes. Fus3p, Kss1p, and Tpk1p showed the greatest occupancy over the transcribed regions of genes. ChIP-Chip experiments show that DNA binding transcription factors and promoter-associated chromatin regulators generally occupy the promoters of genes, whereas transcription elongation factors, gene-associated chromatin regulators, certain histone modifications, and nuclear pore proteins are found enriched along the transcribed regions of genes (figs. S2 and S4). Preferential binding to these factors could explain the localization of the kinases.

Many features of signal transduction pathways are highly conserved in eukaryotes, so it is reasonable to expect that MAPKs and PKAs of higher eukaryotes may also be found to occupy genes that they regulate. Indeed, a human homolog of Hog1p, p38, occupies and activates the myogenin (*MYOG*) and muscle-creatine kinase (*CKM*) promoters during human myogenesis (*10*). The observation that components of many signal transduction pathways physically occupy their target genes upon activation should facilitate the mapping of the regulatory circuitry that eukaryotic cells use to modify gene expression in response to a broad range of environmental cues.

References and Notes

1. C. S. Hill, R. Treisman, *Cell* **80**, 199 (1995).
2. M. Karin, T. Hunter, *Curr. Biol.* **5**, 747 (1995).
3. A. L. Clayton, L. C. Mahadevan, *FEBS Lett.* **546**, 51 (2003).
4. S. H. Yang, A. D. Sharrocks, A. J. Whitmarsh, *Gene* **320**, 3 (2003).
5. J. W. Edmunds, L. C. Mahadevan, *J. Cell Sci.* **117**, 3715 (2004).
6. P. M. Alepuz, A. Jovanovic, V. Reiser, G. Ammerer, *Mol. Cell* **7**, 767 (2001).
7. M. Proft, K. Struhl, *Mol. Cell* **9**, 1307 (2002).
8. P. M. Alepuz, E. de Nadal, M. Zapater, G. Ammerer, F. Posas, *EMBO J.* **22**, 2433 (2003).
9. E. De Nadal *et al.*, *Nature* **427**, 370 (2004).
10. C. Simone *et al.*, *Nat. Genet.* **36**, 738 (2004).
11. W. K. Huh *et al.*, *Nature* **425**, 686 (2003).
12. L. Tomas-Cobos, L. Casadome, G. Mas, P. Sanz, F. Posas, *J. Biol. Chem.* **279**, 22010 (2004).
13. P. Ferrigno, F. Posas, D. Koepp, H. Saito, P. A. Silver, *EMBO J.* **17**, 5606 (1998).
14. S. M. O'Rourke, I. Herskowitz, E. K. O'Shea, *Trends Genet.* **18**, 405 (2002).
15. M. A. Schwartz, H. D. Madhani, *Annu. Rev. Genet.* **38**, 725 (2004).
16. C. J. Roberts *et al.*, *Science* **287**, 873 (2000).
17. S. K. Mahanty, Y. Wang, F. W. Farley, E. A. Elion, *Cell* **98**, 501 (1999).
18. L. Schneper, K. Duvel, J. R. Broach, *Curr. Opin. Microbiol.* **7**, 624 (2004).
19. Y. W. Chang, S. C. Howard, P. K. Herman, *Mol. Cell* **15**, 107 (2004).
20. C. Klein, K. Struhl, *Mol. Cell. Biol.* **14**, 1920 (1994).
21. D. E. Martin, A. Soulard, M. N. Hall, *Cell* **119**, 969 (2004).
22. P. Jorgensen *et al.*, *Genes Dev.* **18**, 2491 (2004).
23. J. M. Casolari *et al.*, *Cell* **117**, 427 (2004).
24. J. M. Casolari, C. R. Brown, D. A. Drubin, O. J. Rando, P. A. Silver, *Genes Dev.* **19**, 1188 (2005).
25. M. Schmid *et al.*, *Mol. Cell* **21**, 379 (2006).
26. F. C. Holstege *et al.*, *Cell* **95**, 717 (1998).
27. D. K. Pokholok *et al.*, *Cell* **122**, 517 (2005).
28. The authors thank G. Fink, M. Guenther, C. Harbison, T. Lee, and S. Levine for critical discussions and S. Levine and E. Herbolsheimer for computational support. The work was supported by NIH grants HG002668 and GM069676. R.A.Y. consults for Agilent Technologies.

Supporting Online Material
www.sciencemag.org/cgi/content/full/313/5786/533/DC1
SOM Text
Figs. S1 to S5
Tables S1 to S7
References

21 March 2006; accepted 24 May 2006
10.1126/science.1127677

Cortical 5-HT$_{2A}$ Receptor Signaling Modulates Anxiety-Like Behaviors in Mice

Noelia V. Weisstaub,[1,3] Mingming Zhou,[2] Alena Lira,[2] Evelyn Lambe,[6]* Javier González-Maeso,[7] Jean-Pierre Hornung,[8] Etienne Sibille,[1]† Mark Underwood,[2] Shigeyoshi Itohara,[9] William T. Dauer,[5] Mark S. Ansorge,[2,3] Emanuela Morelli,[2,3] J. John Mann,[2] Miklos Toth,[10] George Aghajanian,[6] Stuart C. Sealfon,[7] René Hen,[2,4] Jay A. Gingrich[2,3]‡

Serotonin [5-hydroxytryptamine (5-HT)] neurotransmission in the central nervous system modulates depression and anxiety-related behaviors in humans and rodents, but the responsible downstream receptors remain poorly understood. We demonstrate that global disruption of 5-HT2A receptor (5HT2AR) signaling in mice reduces inhibition in conflict anxiety paradigms without affecting fear-conditioned and depression-related behaviors. Selective restoration of 5HT2AR signaling to the cortex normalized conflict anxiety behaviors. These findings indicate a specific role for cortical 5HT2AR function in the modulation of conflict anxiety, consistent with models of cortical, "top-down" influences on risk assessment.

The neurotransmitter serotonin modulates a diverse array of functions related to homeostasis and responses to the environment (*1–11*). Despite the importance of these observations, little is known about the brain structures or the postsynaptic receptors that mediate these effects of 5-HT.

The cortex, ventral striatum, hippocampus, and amygdala are highly enriched in 5HT2AR expression. These structures and their connecting circuits modulate the behavioral response to novelty and threat—behaviors that are typically thought to reflect the anxiety state of the organism (*12*). Given the importance of 5-HT in modulating anxiety states, we sought to determine whether 5HT2AR signaling mediates 5-HT effects on anxiety-related behaviors. We therefore generated genetically modified mice with global disruption of 5HT2AR signaling capacity (*htr2a*$^{-/-}$ mice; fig. S1).

We examined anxiety-related behaviors of *htr2a*$^{-/-}$ mice in several paradigms. The open field (OF) is an arena that presents a conflict between innate drives to explore a novel environment and safety. Under brightly lit conditions, the center of the OF is aversive and potentially risk-laden, whereas exploration of the periphery provides a safer choice. We found that *htr2a*$^{-/-}$ mice explored the center portion of the environment (as a percentage of total exploratory activity) more than their intact *htr2a*$^{+/+}$ littermates did (Fig. 1A; $P < 0.01$). The *htr2a*$^{-/-}$ mice also exhibited more rearing—a maneuver that raises the animal onto its hind limbs, allowing greater visual perspective of the environment but also exposing the animal to greater risk (Fig. 1B; $P < 0.05$).

We examined the behavior of *htr2a*$^{-/-}$ mice in three other conflict paradigms: the dark-light choice test (DLC), the elevated plus-maze (EPM), and the novelty-suppressed feeding (NSF) paradigm. The DLC provides the chance to explore an arena consisting of dark (safe) and brightly lit (risky) areas. The total time of exploratory activity did not differ between genotypes (Fig. 1F); however, *htr2a*$^{-/-}$ mice explored the lit compartment to a greater extent than their *htr2a*$^{+/+}$ littermates, as measured by the percentage of total exploratory time spent in the light compartment (Fig. 1D; $P < 0.05$) and the percentage of total time spent in the light compartment (Fig. 1E; $P < 0.01$). The EPM has two "risk-laden" arms (open without sidewalls) and two "safe" arms (closed by sidewalls). The *htr2a*$^{-/-}$ mice explored the riskier portions of the EPM to a greater extent than the *htr2a*$^{+/+}$ mice, as measured by the percentage of entries made into the open arms (Fig. 1G; $P < 0.05$) and the percentage of time spent in the open arms (Fig. 1H; $P < 0.01$). As in the other tests, total locomotor activity was comparable between genotypes (Fig. 1I). We also examined the effect of *htr2a*$^{-/-}$ mice in the NSF test, which depends less on locomotor activity and is driven by hunger rather than exploratory drive. Consistent with other conflict tests, *htr2a*$^{-/-}$ mice exhibited a shorter latency to begin feeding in a novel environment (Fig. 1J) than the *htr2a*$^{+/+}$ mice ($P < 0.05$), with no differences in feeding activity in the home cage (Fig. 1L) or differences in weight loss (Fig. 1K).

In humans, anxiety and depression often coexist, and altered serotonin signaling has been implicated in the etiology of both disorders (*13*). Therefore, we examined the role of reduced 5HT2AR signaling in depression-related behaviors, as measured by the forced swim test (FST) and the tail suspension test (TST). These paradigms reflect the behavioral response to inescapable stress, not conflict, and are sensitive to antidepressant but not anxiolytic treatments (*14*, *15*). In both tests, rodents usually struggle to escape from these situations, interspersed with periods of immobility that has been interpreted as "behavioral despair" (*16*). When we used these tests to assess *htr2a*$^{-/-}$ mice, we found no difference in immobility when compared to their *htr2a*$^{+/+}$ littermates in either test

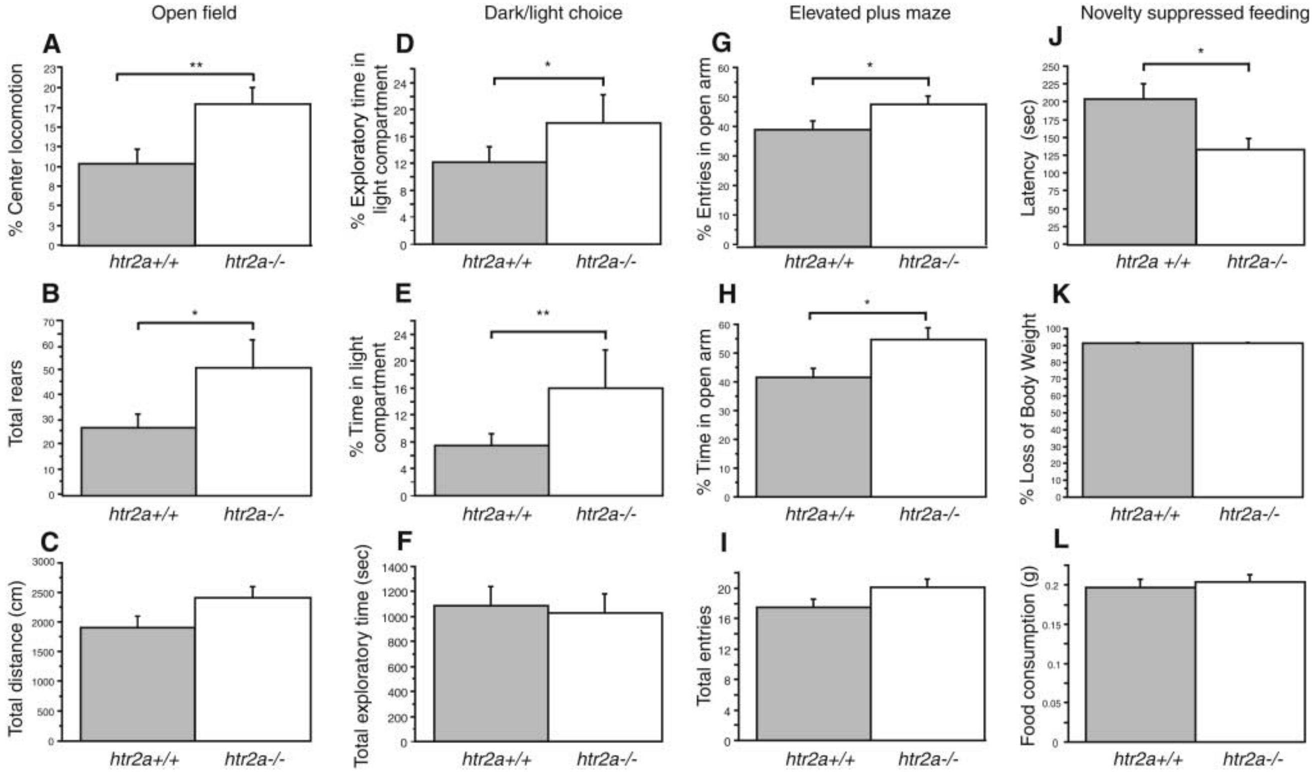


Fig. 1. *htr2a*$^{-/-}$ mice show decreased inhibition in conflict anxiety paradigms. (**A** to **C**) OF measures. (A) percentage of total locomotor activity occurring in the center of the arena. (B) Rearing. (C) Total distance traveled in the periphery and center. (**D** to **F**) DLC measures. (D) Percentage of total exploratory time spent in the light compartment. (E) Percentage of total time spent in the light compartment. (F) Total exploratory time (s). (**G** to **I**) EPM measures. (G) Percentage of total entries made into the open arms. (H) Percentage of time spent in the open arms. (I) Total number of entries into any arm. (**J** to **L**) NSF measures. (J) Latency to approach the food pellet (s). (K) Percentage of body weight lost after deprivation. (L) Amount of food consumed in home cage during 5-min period. *$P < 0.05$; **$P < 0.01$. Mean ± SEM, n = 26 to 39 mice per group.

[1]Department of Biology, [2]Department of Psychiatry, [3]Sackler Institute Laboratories, [4]Center for Neurobiology and Behavior, [5]Department of Neurology, Columbia University and the New York State Psychiatric Institute, New York, NY 10032, USA. [6]Department of Psychiatry, Yale University, New Haven, CT 06520, USA. [7]Department of Neurology, Mount Sinai School of Medicine, New York, NY 10029, USA. [8]Institut de Biologie Cellulaire et de Morphologie, Université de Lausanne, Lausanne, Switzerland. [9]Laboratory for Behavioral Genetics, Riken Brain Science Institute, Wako City, Japan [10]Department of Pharmacology, Cornell University Medical Center, New York, NY 10021, USA.

*Present address: Department of Physiology, University of Toronto, Toronto ON, M5S1A8, Canada.

†Current address: Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA 15213, USA.

‡To whom correspondence should be addressed. E-mail: jag46@columbia.edu

(Fig. 2, A and B). These findings dissociated the low-anxiety phenotype of *htr2a*$^{-/-}$ mice from depression-related behaviors.

To assess the specificity of these findings, we examined other parameters that might influence their outcome. The effect of genotype on exploratory activity was specific to conflict tests because home cage activity did not differ between genotypes. Motor coordination, strength, and sensory processing were unimpaired. We also assessed whether anxiety differences might be due to abnormal hypothalamic-pituitary-adrenal function. Baseline concentrations of plasma corticosterone were comparable in each genotype. Likewise, following novel OF or FST exposure, the rise in corticosterone release was the same in each genotype (fig. S2). We surveyed the content of bioamines and their metabolites in several different brain regions to determine whether the absence of 5HT2AR signaling may have altered the functioning of these systems that are known to influence anxiety-related behaviors. We found no evidence of altered content or turnover of these transmitters as a function of genotype (fig. S5). We assessed the cortical expression of 30 different neurotransmitter receptors using quantitative real-time polymerase chain reaction and found no differences between *htr2a*$^{+/+}$ and *htr2a*$^{-/-}$ mice (with the exception of 5HT2AR expression; table S1).

Although we did not find differences at the mRNA level, differences of receptor expression or coupling might still exist in *htr2a*$^{-/-}$ mice. Because the 5HT$_{2C}$ receptor (5HT2CR) has been implicated in anxiety (*17*), we quantified the amount of agonist-coupled 5HT2CR in *htr2a*$^{+/+}$ and *htr2a*$^{-/-}$ mice using [I^{125}]-DOI [1-(2,5-dimethoxy-4-iodophenyl)-2-aminopropane] autoradiography. No differences in the level of expression of 5HT2CR were observed (fig. S3).

Finally, we also investigated the cellular structure of the cortex, given the high level of expression of 5HT2AR in this brain area. No differences in cell number, mantle thickness, barrel field formation, or the expression of GABA (γ-aminobutyric acid)–containing neuronal markers were seen (fig. S4).

The relation between anxiety and fear is complex because each construct depends on partially overlapping circuitry. Acquisition of fear conditioning requires functional integrity of the hippocampus and the amygdala (*18*), whereas conflict anxiety behaviors implicate the hippocampus, amygdala, cortex, and periaquaductal grey (PAG) (*7*, *19*). To examine whether impaired 5HT2AR signaling in the hippocampus or amygdala disrupts fear-related behaviors, we performed cued and contextual fear-conditioning experiments using an aversive foot-shock stimulus (unconditioned stimulus) paired with a tone (conditioned stimulus). Before the tone-shock pairing, fear-related behavior (i.e., freezing) in the conditioning context was comparable between genotypes (Fig. 2C). After pairing of the conditioning context with the foot shock, we observed increased freezing in response to the context alone with no differences between genotypes (Fig. 2C). When presented with the conditioned tone in an unfamiliar context, mice of both genotypes (previously exposed to paired presentations of tone and foot shock) froze to a greater extent during the tone presentation than during the first minute spent in the new environment (Fig. 2D) and more than control mice previously exposed to unpaired presentation of these stimuli.

The dissociation from learned fear in these studies indicates that the low conflict anxiety shown by *htr2a*$^{-/-}$ mice is not affected by abnormal conditioned fear learning and consequently does not result from altered 5HT2AR signaling in the hippocampus or amygdala. If hippocampal and amygdala functioning is intact, this finding suggests that impaired 5HT2AR signaling in PAG or cortex might underlie their conflict anxiety phenotype. However, the PAG acts to modulate "escape" or freezing behaviors (*20*), which appear to be unaffected in *htr2a*$^{-/-}$ mice. This led us to reason that reduced cortical 5HT2AR signaling may underlie our observed phenotype. We thus attempted to rescue normal conflict behavior in *htr2a*$^{-/-}$ mice by selective restoration of 5HT2AR function to the cortex.

To restore 5HT2AR signaling in the cortex, we capitalized on the methodology used to create our global knockout—namely, an insertion mutation between the promoter and the coding region that blocks transcription and translation of the *htr2a* gene (Fig. 3A and fig. S1). Unidirectional lox-P sites flank the insertion mutation, and under the action of the bacteriophage P1 recombinase, Cre, the inserted sequence can be removed, thus restoring receptor expression under the control of its endogenous promoter (Fig. 3A).

The gene *Emx1* is expressed in the forebrain during early brain maturation (*21*) and has been used to drive *Cre* expression and control forebrain gene expression in other systems (*22*). We crossed *htr2a*$^{-/-}$ mice with mice expressing *Emx1-Cre* to selectively restore 5HT2AR expression to the forebrain while leaving other sites of 5HT2AR expression blocked (*htr2a*$^{-/-}$ × *Emx1-Cre*).

Receptor autoradiography was performed using the agonist [^{125}I]-DOI. In *htr2a*$^{-/-}$ × *Emx1-Cre* mice, we observed that 5HT2AR expression was restored principally in layer V of the cortex and in a closely associated structure, the claustrum (*23*). No measurable expression was seen in the hippocampus, a structure expressing *Emx1*. We found no significant 5HT2AR mRNA expression in the striatum of *htr2a*$^{-/-}$ × *Emx1-Cre* mice as compared to *htr2a*$^{-/-}$ mice (fig S6A). Likewise, the thalamus and other subcortical structures that express 5HT2AR, but not *Emx1*, were devoid of expression (Fig. 3C).

To determine whether compensatory alterations in 5HT2CR expression were present in *htr2a*$^{-/-}$ mice or *htr2a*$^{-/-}$ × *Emx1-Cre* mice, we assessed 5HT2CR mRNA expression (fig. S6B). We found no evidence of 5HT2CR alterations in *htr2a*$^{-/-}$ × *Emx1-Cre* mice.

To verify the functionality of the restored cortical 5HT2AR, we assessed the electrophysiological response of cortical slices to 5-HT. We performed whole-cell recordings of layer V pyramidal neurons in cortical slices from *htr2a*$^{+/+}$, *htr2a*$^{-/-}$, and *htr2a*$^{-/-}$ × *Emx1-Cre* mice. There were no significant differences among these groups in resting potential, input resistance, and spike amplitude. However, 5-HT

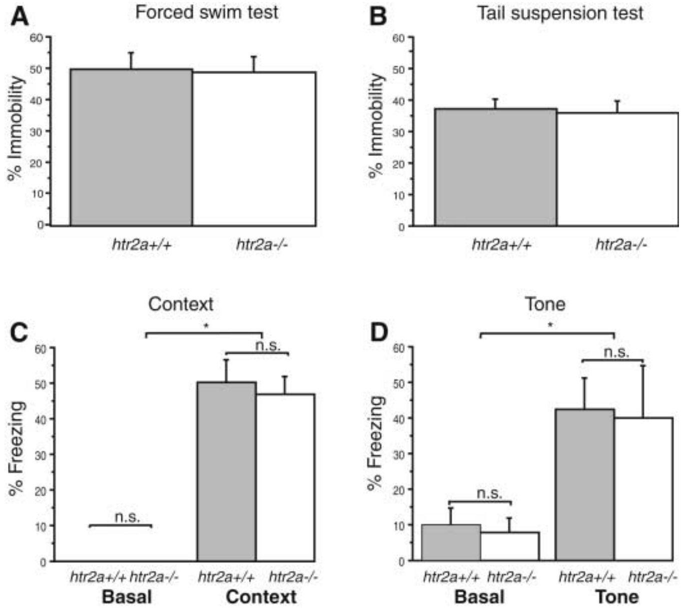


Fig. 2. Depression and fear-related measures are not affected in *htr2a*$^{-/-}$ mice. (**A**) FST: Percentage of time spent immobile during the 4-min test. (**B**) TST: Percentage of time spent immobile during the 7-min test. (**C** and **D**) Fear-conditioned learning. (C) Mean percentage of freezing in basal condition measured during the first 60 s in the first day of exposure and mean percentage of freezing time during context test. (D) Percentage of freezing time in new context without and during the presence of the cue test. Mean ± SEM, $n = 12$ to 40 mice per group for all tests.

produced robust increases in spontaneous excitatory postsynaptic currents (sEPSCs) in pyramidal neurons from *htr2a*$^{+/+}$ and *htr2a*$^{-/-}$ × *Emx1-Cre* mice, but not in *htr2a*$^{-/-}$ mice (Fig. 3D; $P < 0.0001$). The selective 5HT2AR antagonist, MDL 100907, blocked the 5-HT–elicited increases in sEPSC frequency, but had no effect in *htr2a*$^{-/-}$ mice. Norepinephrine (NE) increased sEPSCs to an equal extent in all

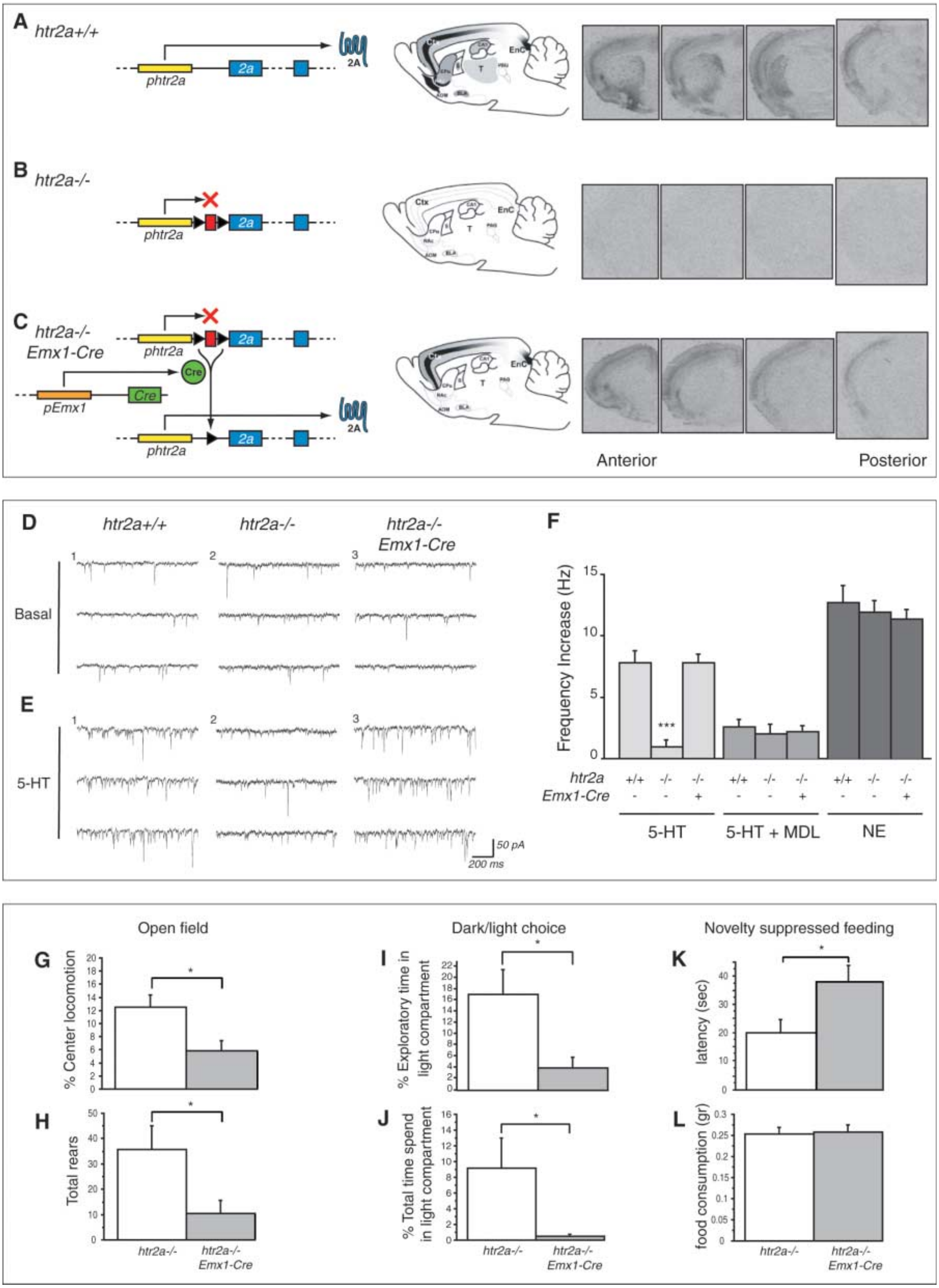


Fig. 3. Cortical restoration of 5HT2AR function normalizes conflict anxiety in *htr2a*$^{-/-}$ mice. (**A** to **C**) Filled blue boxes represent exons of *htr2a* gene. Narrow boxes labeled with p*htr2a* or p*Emx1* represent the endogenous promoters for each gene. Serpentine symbol indicates the *htr2a* gene product. (Left) (A) Schematic of the wild-type *htr2a* locus. (B) Lox-p (triangles)–flanked cassette (red box) inserted upstream from the first initiation codon of the *htr2a* gene blocks transcription and translation. (C) Expression of Cre under the control of the *Emx1* promoter interacts with the lox-p sequences to remove the cassette and restore expression of *htr2a* gene. (Middle) Schematic representation of the pattern of expression of 5HT2AR in *htr2a*$^{+/+}$ (A), *htr2a*$^{-/-}$ (B), and *htr2a*$^{-/-}$ × *Emx1-Cre* (C) mice. Abbreviations: CTX, cortex; T, thalamus; CA1, CA1 region of hippocampus; PAG, periaquaductal grey; CPu, caudate-putamen; NAc, nucleus accumbens; BLA, basolateral nucleus amygdala; AOM, anterior olfactory nucleus (medial); EnC, entorhinal cortex. (Right) Autoradiography with [^{125}I]-DOI in *htr2a* $^{+/+}$ (A), *htr2a* $^{-/-}$ (B), *htr2a*$^{-/-}$ × *Emx1-Cre* (C) mice shown at representative anterior and posterior slices. (**D**) Voltage-clamp recordings under basal conditions from (1) *htr2a*$^{+/+}$, (2) *htr2a*$^{-/-}$, and (3) *htr2a*$^{-/-}$ × *Emx1-Cre* mice. (**E**) Voltage-clamp recordings of the peak response to bath-applied 5-HT (100 μM, 1 min) in the same neurons. (**F**) Bar graph showing changes in sEPSC frequency in neurons from *htr2a*$^{+/+}$ and *htr2a*$^{-/-}$ × *Emx1-Cre* mice, using 5-HT (100 μM), 5-HT (100 μM) + MDL 100907 (100 nM), and NE (100 μM). (**G** and **H**) OF measures. (**I** and **J**) DLC measures. (**K** and **L**) NSF. See Fig. 1 for explanations. *$P < 0.05$, ***$P < 0.0001$. Mean ± SEM, n = 10 to 12 neurons per genotype, n = 13 to 14 mice per group for behavioral experiments.

groups, indicating that the loss of 5HT2AR signaling had no effect on the response to other bioamines (Fig. 3E).

To determine whether restored cortical 5HT2AR signaling was sufficient to normalize conflict behavior, we used three paradigms that previously elicited a robust phenotype in *htr2a*$^{-/-}$ mice: OF, DLC, and NSF. In the OF, mice with cortical restoration of 5HT2AR signaling exhibited wild-type levels of anxiety-like behavior as measured by the percentage of exploratory activity in the center of the field (Fig. 3G; $P < 0.05$) and rearing (Fig. 3H; $P < 0.05$). Similar effects of the cortical 5HT2AR rescue on anxiety were seen in the DLC [decreased percentage of exploratory time (Fig. 3I; $P < 0.05$) and decreased percentage of total time (Fig. 3J; $P < 0.05$) in the light compartment as compared to *htr2a*$^{-/-}$ mice] and the NSF (increased latency; Fig. 3K, $P < 0.05$ compared to *htr2a*$^{-/-}$ mice). Corroborating the specificity of these anxiety-related findings, behavioral responses in depression-related paradigms, such as the FST and TST, were unchanged in *htr2a*$^{-/-}$ × *Emx1-Cre* mice (fig. S7) as compared with *htr2a* $^{-/-}$ littermates. A similar strategy when used to restore 5HT2AR expression to a subcortical region (i.e., thalamus) produced no difference between rescue and *htr2a*$^{-/-}$ mice in the DLC (see supporting online material), supporting the specificity of the cortex in the normalization of anxiety-related behaviors.

The tissue-specific restoration of an endogenous gene product to a knockout animal provides a precise method for assessing the role of specific circuits in modulating behavior. In addition, when a tissue-restricted rescue normalizes the lost function of a global knockout, such a finding offsets many of the interpretive problems that arise with loss-of-function mutations. In our study, the absence of measurable adaptations in the *htr2a*$^{-/-}$ mice, combined with the reversal of their phenotype by a selective reactivation of *htr2A* gene expression in the cortex, suggests that nonspecific developmental alterations are unlikely to explain our findings.

The precise role of 5-HT signaling in anxiety appears to be complex. Mice with mutations of the 5-HT plasma membrane transporter or 5-HT$_{1A}$ receptor exhibit elevated anxiety levels, but the effects of these mutations on anxiety have been attributed to altered brain development (*24*, *25*). In contrast, the low-anxiety phenotype of *htr2a*$^{-/-}$ mice does not appear to be related to altered brain development, but it may be related to the chronic nature of the mutation in the adult mice. Attempts to reduce conflict anxiety with acute pharmacological administration of 5HT2AR antagonists have been unsuccessful (*26*) or mixed (*27*), whereas chronic reduction of 5HT2AR signaling through the use of antisense receptor down-regulation methods has proven quite effective (*28*). The need for chronic blockade or down-regulation of 5HT2ARs is consistent with the properties of serotonergic anxiolytics that require several weeks to achieve therapeutic effects.

The cortex has been hypothesized as a "top-down" modulator of anxiety-related processes, given the extensive interconnections between the cortex and structures such as the hippocampus and amygdala. Recent functional imaging data in human subjects support this notion (*29*–*31*). Thus, it is intriguing that 5-HT signaling in the cortex can exert pronounced effects on behavior in conflict anxiety tests. A primary role of cortical 5HT2AR signaling in risk or threat assessment may explain the specificity of *htr2a* disruption on conflict anxiety and the absence of effects on conditioned fear and depression-related behaviors. Indeed, modulation of layer V pyramidal neuron glutamate release by 5HT2AR signaling is a likely mechanism by which these cortical projection neurons could modify the activity of subcortical structures. Given the complex effects of 5-HT on a variety of central nervous system functions, a better understanding of the receptor and neural substrates that mediate them may lead to a more nuanced view of 5-HT function and improved therapeutics for anxiety and affective disorders.

References and Notes

1. I. Artaiz, A. Zazpe, J. Del Rio, *Behav. Pharmacol.* **9**, 103 (1998).
2. J. C. Ballenger, F. K. Goodwin, L. F. Major, G. L. Brown, *Arch. Gen. Psychiatry* **36**, 224 (1979).
3. P. Bjorntorp, *Metabolism* **44**, 38 (1995).
4. L. Booij, A. J. Van der Does, W. J. Riedel, *Mol. Psychiatry* **8**, 951 (2003).
5. P. K. Bridges, J. R. Bartlett, P. Sepping, B. D. Kantamaneni, G. Curzon, *Psychol. Med.* **6**, 399 (1976).
6. F. K. Goodwin, R. M. Post, *Br. J. Clin. Pharmacol.* **15** (Suppl. 3), 393S (1983).
7. F. G. Graeff, *Psychopharmacology (Berlin)* **163**, 467 (2002).
8. S. L. Handley, J. W. McBlane, M. A. Critchley, K. Njung'e, *Behav. Brain Res.* **58**, 203 (1993).
9. M. Luciana, E. D. Burgund, M. Berman, K. L. Hanson, *J. Psychopharmacol.* **15**, 219 (2001).
10. R. A. McFarlain, J. M. Bloom, *Psychopharmacologia* **27**, 85 (1972).
11. L. H. Tecott *et al.*, *Nature* **374**, 542 (1995).
12. M. J. Millan, *Prog. Neurobiol.* **70**, 83 (2003).
13. K. J. Ressler, C. B. Nemeroff, *Depress. Anxiety* **12** (Suppl. 1), 2 (2000).
14. R. D. Porsolt, G. Anton, N. Blavet, M. Jalfre, *Eur. J. Pharmacol.* **47**, 379 (1978).
15. L. Steru, R. Chermat, B. Thierry, P. Simon, *Psychopharmacology (Berlin)* **85**, 367 (1985).
16. F. Bai, X. Li, M. Clay, T. Lindstrom, P. Skolnick, *Pharmacol. Biochem. Behav.* **70**, 187 (2001).
17. M. D. Wood, *Curr. Drug Targets CNS Neurol. Disord.* **2**, 383 (2003).
18. N. R. Selden, B. J. Everitt, L. E. Jarrard, T. W. Robbins, *Neuroscience* **42**, 335 (1991).
19. S. Shibata, K. Yamashita, E. Yamamoto, T. Ozaki, S. Ueki, *Psychopharmacology (Berlin)* **98**, 38 (1989).
20. V. de Paula Soares, H. Zangrossi Jr., *Brain Res. Bull.* **64**, 181 (2004).
21. E. Boncinelli, M. Gulisano, V. Broccoli, *J. Neurobiol.* **24**, 1356 (1993).
22. T. Iwasato *et al.*, *Nature* **406**, 726 (2000).
23. T. Miyashita, S. Nishimura-Akiyoshi, S. Itohara, K. S. Rockland, *Neuroscience* **136**, 487 (2005).
24. C. Gross *et al.*, *Nature* **416**, 396 (2002).
25. M. S. Ansorge, M. Zhou, A. Lira, R. Hen, J. A. Gingrich, *Science* **306**, 879 (2004).
26. G. Griebel, G. Perrault, D. J. Sanger, *Neuropharmacology* **36**, 793 (1997).
27. P. O. Mora, C. F. Netto, F. G. Graeff, *Pharmacol. Biochem. Behav.* **58**, 1051 (1997).
28. H. Cohen, *Depress. Anxiety* **22**, 84 (2005).
29. J. M. Kent *et al.*, *Am. J. Psychiatry* **162**, 1379 (2005).
30. S. Bishop, J. Duncan, M. Brett, A. D. Lawrence, *Nat. Neurosci.* **7**, 184 (2004).
31. A. Heinz *et al.*, *Nat. Neurosci.* **8**, 20 (2005).

32. We thank P. Svenningson for help with 5HT2AR in situ hybridization experiments, G. Marek for help with early electrophysiology experiments, H. Westphal for sharing the "stop" cassette, Y. Huang for help with bioamine analysis, and F. Menzaghi and M. Milekic for critical reading of the manuscript. Funding sources include National Institute of Mental Health grant KO8 MH01711 (J.A.G.), National Institute on Drug Abuse grant P01 DA12923 (J.A.G., R.H., and S.C.S.), the Whitehall Foundation, the American Foundation for Suicide Prevention, the Gatsby Foundation, the Lieber Center for Schizophrenia Research at Columbia University, and the National Alliance for Research on Schizophrenia and Depression Foundation (J.A.G., E.L., M.S.A.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/313/5786/536/DC1
Materials and Methods
Figs. S1 to S7
References

5 December 2005; accepted 8 June 2006
10.1126/science.1123432